(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57277 A2

(51) International Patent Classification[7]: C12Q 1/68, C07K 14/47, C07H 21/04

(21) International Application Number: PCT/US01/00669

(22) International Filing Date: 30 January 2001 (30.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/180,312	4 February 2000 (04.02.2000)	US
60/207,456	26 May 2000 (26.05.2000)	US
09/608,408	30 June 2000 (30.06.2000)	US
09/632,366	3 August 2000 (03.08.2000)	US
60/234,687	21 September 2000 (21.09.2000)	US
60/236,359	27 September 2000 (27.09.2000)	US
0024263.6	4 October 2000 (04.10.2000)	GB

(71) Applicant *(for all designated States except US)*: **MOLECULAR DYNAMICS, INC.** [—/US]; 928 East Arques Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PENN, Sharron, G.** [GB/US]; 617 South Delaware Street, San Mateo, CA 94402 (US). **HANZEL, David, K.** [US/US]; 968 Loma Verde Avenue, Palo Alto, CA 94303 (US). **CHEN, Wensheng** [CN/US]; 210 Easy Street #25, Mountain View, CA 94043 (US). **RANK, David, R.** [US/US]; 117 El Dorado Commons, Fremont, CA 94539 (US).

(74) Agent: **RONNING, Royal, N., Jr.**; Amersham Pharmacia Biotech, Inc., 800 Centennial Avenue, Piscataway, NJ 08855 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN FETAL LIVER

(57) Abstract: A single exon nucleic acid microarray comprising a plurality of single exon nucleic acid probes for measuring gene expression in a sample derived from human Fetal liver is described. Also described are single exon nucleic acid probes expressed in the Fetal liver and their use in methods for detecting gene expression.

WO 01/57277 A2

HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN FETAL LIVER

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. patent application serial nos. 09/632,366, filed August 3, 2000 and 09/608,408, filed June 30, 2000; claims the benefit under 35 U.S.C. s 119(e) of U.S.provisional patent application serial nos. 60/236,359, filed September 27, 2000, 60/234,687, filed September 21, 2000, 60/207,456, filed May 26, 2000, and 60/180,312, filed February 4, 2000; and further claims the benefit under 35 U.S.C. s 119(a) of UK patent application no. 0024263.6, filed October 4, 2000, the disclosures of which are incorporated herein by reference in their entireties.

REFERENCE TO SEQUENCE LISTING AND INCORPORATION BY REFERENCE THEREOF

The present application includes a Sequence Listing in electronic format, filed pursuant to PCT Administrative Instructions 801 - 806 on a single CD-R disc, in triplicate, containing a file named pto_FETAL_LIVER.txt, created 24 January 2001, having 25,630,231 bytes. The Sequence Listing contained in said file on said disc is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to genome-derived single exon microarrays useful for verifying the expression of regions of genomic DNA predicted to encode protein. In particular, the present invention relates to unique genome-derived single exon nucleic acid probes expressed in human

Fetal liver and single exon nucleic acid microarrays that include such probes.

Background of the Invention

For almost two decades following the invention of general techniques for nucleic acid sequencing, Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 70(4):1209-13 (1973); Gilbert *et al.*, *Proc. Natl. Acad. Sci. USA* 70(12):3581-4 (1973), these techniques were used principally as tools to further the understanding of proteins – known or suspected – about which a basic foundation of biological knowledge had already been built. In many cases, the cloning effort that preceded sequence identification had been both informed and directed by that antecedent biological understanding.

For example, the cloning of the T cell receptor for antigen was predicated upon its known or suspected cell type-specific expression, by its suspected membrane association, and by the predicted assembly of its gene via T cell-specific somatic recombination. Subsequent sequencing efforts at once confirmed and extended understanding of this family of proteins. Hedrick *et al.*, *Nature* 308(5955):153-8 (1984).

More recently, however, the development of high throughput sequencing methods and devices, in concert with large public and private undertakings to sequence the human and other genomes, has altered this investigational paradigm: today, sequence information often precedes understanding of the basic biology of the encoded protein product.

One of the approaches to large-scale sequencing is predicated upon the proposition that expressed sequences – that is, those accessible through isolation of mRNA – are of greatest initial interest. This "expressed sequence tag" ("EST") approach has already yielded vast

amounts of sequence data (see for example Adams *et al.*, *Science* 252:1651 (1991); Williamson, *Drug Discov. Today* 4:115 (1999)). For nucleic acids sequenced by this approach, often the only biological information that is known *a priori* with any certainty is the likelihood of biologic expression itself. By virtue of the species and tissue from which the mRNA had originally been obtained, most such sequences are also annotated with the identity of the species and at least one tissue in which expression appears likely.

More recently, the pace of genomic sequencing has accelerated dramatically. When genomic DNA serves as the initial substrate for sequencing efforts, expression cannot be presumed; often the only *a priori* biological information about the sequence includes the species and chromosome (and perhaps chromosomal map location) of origin.

With the ever-accelerating pace of sequence accumulation by directed, EST, and genomic sequencing approaches – and in particular, with the accumulation of sequence information from multiple genera, from multiple species within genera, and from multiple individuals within a species – there is an increasing need for methods that rapidly and effectively permit the functions of nucleic sequences to be elucidated. And as such functional information accumulates, there is a further need for methods of storing such functional information in meaningful and useful relationship to the sequence itself; that is, there is an increasing need for means and apparatus for annotating raw sequence data with known or predicted functional information.

Although the increase in the pace of genomic sequencing is due in large part to technological changes in sequencing strategies and instrumentation, Service, *Science* 280:995 (1998); Pennisi, *Science* 283: 1822-1823 (1999), there is an important functional motivation as well.

While it was understood that the EST approach would rarely be able to yield sequence information about the noncoding portions of the genome, it now also appears the EST approach is capable of capturing only a fraction of a genome's actual expression complexity.

For example, when the *C. elegans* genome was fully sequenced, gene prediction algorithms identified over 19,000 potential genes, of which only 7,000 had been found by EST sequencing. *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998). Analogously, the recently completed sequence of chromosome 2 of *Arabidopsis* predicts over 4000 genes, Lin *et al.*, *Nature*, 402:761 (1999), of which only about 6% had previously been identified via EST sequencing efforts. Although the human genome has the greatest depth of EST coverage, it is still woefully short of surrendering all of its genes. One recent estimate suggests that the human genome contains more than 146,000 genes, which would at this point leave greater than half of the genes undiscovered. It is now predicted that many genes, perhaps 20 to 50%, will only be found by genomic sequencing.

There is, therefore, a need for methods that permit the functional regions of genomic sequence – and most importantly, but not exclusively, regions that function to encode genes – to be identified.

Much of the coding sequence of the human genome is not homologous to known genes, making detection of open reading frames ("ORFs") and predictions of gene function difficult. Computational methods exist for predicting coding regions in eukaryotic genomes. Gene prediction programs such as GRAIL and GRAIL II, Uberbacher *et al.*, *Proc. Natl. Acad. Sci. USA* 88(24):11261-5 (1991); Xu *et al.*, *Genet. Eng.* 16:241-53 (1994); Uberbacher *et al.*, *Methods Enzymol.* 266:259-81 (1996); GENEFINDER, Solovyev *et al.*, *Nucl. Acids. Res.* 22:5156-63 (1994); Solovyev *et al.*,

Ismb 5:294-302 (1997); and GENESCAN, Burge *et al.*, *J. Mol. Biol.* 268:78-94 (1997), predict many putative genes without known homology or function. Such programs are known, however, to give high false positive rates. Burset *et al.*, *Genomics* 34:353-367 (1996). Using a consensus obtained by a plurality of such programs is known to increase the reliability of calling exons from genomic sequence. Ansari-Lari *et al.*, *Genome Res.* 8(1):29-40 (1998)

Identification of functional genes from genomic data remains, however, an imperfect art. For example, in reporting the full sequence of human chromosome 21, the Chromosome 21 Mapping and Sequencing Consortium reports that prior bioinformatic estimates of human gene number may need to be revised substantially downwards. *Nature* 405:311-199 (2000); Reeves, *Nature* 405:283-284 (2000).

Thus, there is a need for methods and apparatus that permit the functions of the regions identified bioinformatically – and specifically, that permit the expression of regions predicted to encode protein – readily to be confirmed experimentally.

Recently, the development of nucleic acid microarrays has made possible the automated and highly parallel measurement of gene expression. *Reviewed in* Schena (ed.), DNA Microarrays : A Practical Approach (Practical Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376).

It is common for microarrays to be derived from cDNA/EST libraries, either from those previously described in the literature, such as those from the I.M.A.G.E. consortium, Lennon *et al.*, *Genomics* 33(1):151-2 (1996), or from the construction of "problem specific" libraries targeted at a particular biological question, R.S. Thomas

et al., *Cancer Res.* (in press). Such microarrays by definition can measure expression only of those genes found in EST libraries, and thus have not been useful as probes for genes discovered solely by genomic sequencing.

The utility of using whole genome nucleic acid microarrays to answer certain biological questions has been demonstrated for the yeast *Saccharomyces cerevisiae*. De Risi *et al.*, *Science* 278:680 (1997). The vast majority of yeast nuclear genes, approximately 95% however, are single exon genes, *i.e.*, lack introns, Lopez *et al.*, *RNA* 5:1135-1137 (1999); Goffeau *et al.*, *Science* 274:563-67 (1996), permitting coding regions more readily to be identified. Whole genome nucleic acid microarrays have not generally been used to probe gene expression from more complex eukaryotic genomes, and in particular from those averaging more than one intron per gene.

Diseases of the liver are a significant cause of human morbidity and mortality. Increasingly, genetic factors are being found that contribute to predisposition, onset, and/or aggressiveness of most, if not all, of these diseases; although causative mutations in single genes have been identified for some, these disorders are believed for the most part to have polygenic etiologies. There is a need for methods and apparatus that permit prediction of diseases of the liver, particularly those diseases with polygenic etiology, from diagnosis of fetal liver.

Summary of the Invention

The present invention solves these and other problems in the art by providing methods and apparatus for predicting, confirming, and displaying functional information derived from genomic sequence. The present invention also provides apparatus for verifying the expression of putative genes identified within genomic

sequence.

In particular, the invention provides novel genome-derived single exon nucleic acid microarrays useful for verifying the expression of putative genes identified within genomic sequence.

The present invention also provides compositions and kits for the ready production of nucleic acids identical in sequence to, or substantially identical in sequence to, probes on the genome-derived single exon microarrays of the present invention.

Accordingly, in a first aspect of the invention, there is provided a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human Fetal liver, comprising a plurality of single exon nucleic acid probes according to any one of the nucleotide sequences set out in SEQ ID NOs: 1 - 12,673 or a complementary sequence, or a portion of such a sequence.

By plurality is meant at least two, suitably at least 20, most suitably at least 100, preferably at least 1000 and, most preferably, upto 5000.

In one embodiment of the first aspect, each of said plurality of probes is separately and addressably amplifiable.

In an alternative embodiment, each of said plurality of probes is separately and addressably isolatable from said plurality.

In a preferred embodiment, each of said plurality of probes is amplifiable using at least one common primer. Preferably, each of said plurality of probes is amplifiable using a first and a second common primer.

In yet another embodiment, said set of single exon nucleic acid probes comprises between 50 - 20,000 probes, for example, 50 - 5000.

Suitably, said set of single exon nucleic acid

probes comprises at least 50 - 1000 discrete single exon nucleic acid probes having a sequence as set out in any of SEQ ID NOS.: 1 - 25,129 or a complimentary sequence, or a portion of such a sequence.

Preferably, the average length of the single exon nucleic acid probes is between 200 and 500 bp. It is preferred that the average length should be at least 200bp, suitably at least 250bp, most suitably at least 300bp, preferably at least 400bp and, most preferably, 500 bp.

In another embodiment, the single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence. It is preferred that at least 50%, suitably at least 60%, most suitably at least 70%, preferably at least 75%, more preferably at least 80, 85, 90, 95 or 99% of said single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence.

In another preferred embodiment, said single exon nucleic acid lack homopolymeric stretches of A or T. It is preferred that at least 50%, suitably at least 60%, most suitably at least 70%, preferably at least 75%, more preferably at least 80, 85, 90, 95 or 99% of said single exon nucleic acid probes lack homopolymeric stretches of A or T.

Preferably, a spatially-addressable set of single exon nucleic acid probes in accordance with the first aspect of the invention is is addressably disposed upon a substrate.

Suitable substrates include a filter membrane which may, preferably, be nitrocellulose or nylon. The nylon may preferably, be positively-charged. Other suitable substrates include glass, amorphous silicon, crystalline silicon, and plastic. Further suitable materials include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate,

polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, and mixtures thereof.

In a second aspect of the invention, there is provided a microarray comprising a spatially addressable set of single exon nucleic acid probes in accordance with the first aspect of the invention.

In one embodiment, a genome-derived single-exon microarray is packaged together with such an ordered set of amplifiable probes corresponding to the probes, or one or more subsets of probes, thereon. In alternative embodiments, the ordered set of amplifiable probes is packaged separately from the genome-derived single exon microarray.

In another aspect, the invention provides genome-derived single exon nucleic acid probes useful for gene expression analysis, and particularly for gene expression analysis by microarray. In particular embodiments of this aspect, the present invention provides human single-exon probes that include specifically-hybridizable fragments of SEQ ID Nos. 12,674 - 25,129, wherein the fragment hybridizes at high stringency to an expressed human gene. In particular embodiments, the invention provides single exon probes comprising SEQ ID Nos. 1 - 12,673.

Accordingly, in a third aspect of the invention, there is provided a single exon nucleic acid probe for measuring human gene expression in a sample derived from human Fetal liver which is a nucleic acid molecule comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 1 - 12,673 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human Fetal liver.

In one embodiment, a single exon nucleic acid probe in accordance with the third aspect comprises a nucleotide sequence as set out in any of SEQ ID NOs.:

12,674 - 25,129 or a complementary sequence or a fragment thereof.

In a fourth aspect of the invention, there is provided a single exon nucleic acid probe for measuring human gene expression in a sample derived from human Fetal liver which is a nucleic acid molecule having a sequence encoding a peptide comprising a peptide sequence as set out in any of SEQ ID NOs.: 25,130 - 37,156 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human Fetal liver.

Preferably, a single exon nucleic acid probe in accordance with the third or fourth aspects of the invention comprises between at least 15 and 50 contiguous nucleotides of said SEQ ID NO:. It is preferred that the single exon nucleic acid probe comprises at least 15, suitably at least 20, more suitably at least 25 or preferably at least 50 contiguous nucleotides of said SEQ ID NO:.

In another preferred embodiment, a single exon nucleic acid probe in accordance with the third or fourth aspects of the invention is between 3kb and 25kb in length. It is preferred that said probe is no more than 3kb, suitably no more than 5kb, more suitably no more than 10kb, preferably 15kb, more preferably 20kb or, most preferably, no more than 20kb in length.

Preferably, a single exon nucleic acid probe in accordance with either the fifth or sixth aspect of the invention is DNA, preferably single-stranded DNA, RNA or PNA.

In another embodiment of either the third or fourth aspect of the invention, a single exon nucleic acid probe is detectably labeled. Suitable detectable labels include a radionuclide, a fluorescent label or a first member of a specific binding pair. Suitable fluorescent

labels include dyes such as cyanine dyes, preferably Cy3 and Cy5 although other suitable dyes will be known to those skilled in the art.

In a particularly preferred embodiment, a single exon nucleic acid probe in accordance with either the third or fourth aspect of the invention lacks prokaryotic and bacteriophage vector sequence. In yet another embodiment, a single exon nucleic acid probe in accordance with either the third or fourth aspect of the invention lacks homopolymeric stretches of A or T.

In a fifth aspect of the invention, there is provided an amplifiable nucleic acid composition, comprising:

the single exon nucleic acid probe in accordance with either of the third or fourth aspects of the invention; and at least one nucleic acid primer;

wherein said at least one primer is sufficient to prime enzymatic amplification of said probe.

In an sixth aspect of the invention, there is provided a method of measuring gene expression in a sample derived from human Fetal liver, comprising:

contacting the single exon microarray in accordance with the second aspect of the invention, with a first collection of detectably labeled nucleic acids, said first collection of nucleic acids derived from mRNA of human Fetal liver; and then

measuring the label detectably bound to each probe of said microarray.

In a seventh aspect of the invention, there is provided a method of identifying exons in a eukaryotic genome, comprising:

algorithmically predicting at least one exon from genomic sequence of said eukaryote; and then

detecting specific hybridization of detectably labeled nucleic acids to a single exon probe,

wherein said detectably labeled nucleic acids are derived from mRNA from the Fetal liver of said eukaryote, said probe is a single exon probe having a fragment identical in sequence to, or complementary in sequence to, said predicted exon, said probe is included within a single exon microarray in accordance with the first aspect of the invention, and said fragment is selectively hybridizable at high stringency.

In a eighth aspect of the invention, there is provided a method of assigning exons to a single gene, comprising:

identifying a plurality of exons from genomic sequence in accordance with the seventh aspect of the invention; and then

measuring the expression of each of said exons in a plurality of tissues and/or cell types using hybridization to single exon microarrays having a probe with said exon,

wherein a common pattern of expression of said exons in said plurality of tissues and/or cell types indicates that the exons should be assigned to a single gene.

In an ninth aspect of the invention, there is provided a nucleic acid sequence as set out in any of SEQ ID NOs: 1 - 25,129 wherein said sequence encodes a peptide.

In a tenth aspect of the invention, there is provided a peptide encoded by a sequence comprising a sequence as set out in any of SEQ ID NOs: 12,674 - 25,129, or a complementary sequence or coding portion thereof.

In a preferred embodiment, a peptide may be encoded by a sequence comprising a sequence set out in any of SEQ ID NOS.: 1 -12,673.

In a further aspect, the invention provides peptides comprising an amino acid sequence translated from the DNA fragments, said amino acid sequences comprising SEQ

ID NOS.: 25,130 - 37,156.

Accordingly in a eleventh aspect of the invention there is provided a peptide comprising a sequence as set out in any of SEQ ID NOs: 25,130 - 37,156, or fragment thereof.

In another aspect, the invention provides means for displaying annotated sequence, and in particular, for displaying sequence annotated according to the methods and apparatus of the present invention. Further, such display can be used as a preferred graphical user interface for electronic search, query, and analysis of such annotated sequence.

Detailed Description of the Invention

Definitions

As used herein, the term "microarray" and phrase "nucleic acid microarray" refer to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed.

As so defined, the term "microarray" and phrase "nucleic acid microarray" include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); and Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376). As so defined, the term "microarray" and phrase "nucleic acid microarray" further include substrate-bound collections of plural nucleic acids in which the nucleic acids are distributably disposed on a plurality of beads, rather than on a unitary

planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000); in such case, the term "microarray" and phrase "nucleic acid microarray" refer to the plurality of beads in aggregate.

As used herein with respect to a nucleic acid microarray, the term "probe" refers to the nucleic acid that is, or is intended to be, bound to the substrate; in such context, the term "target" thus refers to nucleic acid intended to be bound thereto by Watson-Crick complementarity. As used herein with respect to solution phase hybridization, the term "probe" refers to the nucleic acid of known sequence that is detectably labeled.

As used herein, the expression "probe comprising SEQ ID NO.", and variants thereof, intends a nucleic acid probe, at least a portion of which probe has either (i) the sequence directly as given in the referenced SEQ ID NO., or (ii) a sequence complementary to the sequence as given in the referenced SEQ ID NO., the choice as between sequence directly as given and complement thereof dictated by the requirement that the probe hybridize to mRNA.

As used herein, the term "open reading frame" and the equivalent acronym "ORF" refer to that portion of an exon that can be translated in its entirety into a sequence of contiguous amino acids i.e. a nucleic acid sequence that, in at least one reading frame, does not possess stop codons; the term does not require that the ORF encode the entirety of a natural protein.

As used herein, the term "amplicon" refers to a PCR product amplified from human genomic DNA, containing the predicted exon.

As used herein the term "exon" refers to the consensus prediction of the various exon and gene predicting algorithms i.e. a nucleic acid sequence bioinformatically predicted to encode a portion of a

natural protein.

As used herein, the term "peptide" refers to a sequence of amino acids. The sequences referred to as PEPTIDE SEQ ID NOS.: are the predicted peptide sequences that would be translated from one of the exons, or a portion thereof set out in exon SEQ ID NOS.:. The codons encoding the peptide are wholly contained within the exon.

As used herein, a "portions" of a defined nucleotide sequence or sequences can be and, preferably, are fragments unique to that sequence or to one or a combination of those sequences. A fragment unique to a nucleic acid molecule is one that is a signature for the larger nucleic acid molecule.

As used herein, the phrase "expression of a probe" and its linguistic variants means that the ORF present within the probe, or its complement, is present within a target mRNA.

As used herein, "stringent conditions" refers to parameters well known to those skilled in the art. When a nucleic acid molecule is said to be hybridisable to another of a given sequence under "stringent conditions" it is meant that it is homologous to the given sequence.

As used herein, the phrase "specific binding pair" intends a pair of molecules that bind to one another with high specificity. Binding pairs are said to exhibit specific binding when they exhibit avidity of at least 10^7, preferably at least 10^8, more preferably at least 10^9 liters/mole. Nonlimiting examples of specific binding pairs are: antibody and antigen; biotin and avidin; and biotin and streptavidin.

As used herein with respect to the visual display of annotated genomic sequence, the term "rectangle" means any geometric shape that has at least a first and a second border, wherein the first and second borders each are capable of mapping uniquely to a point of another visual

object of the display.

As used herein, a "Mondrian" means a visual display in which a single genomic sequence is annotated with predicted and experimentally confirmed functional information.

Brief Description of the Drawings

The present invention is further illustrated with reference to the following non-limiting figures and examples in which:

FIG. 1 illustrates a process for predicting functional regions from genomic sequence, confirming the functional activity of such regions experimentally, and associating and displaying the data so obtained in meaningful and useful relationship to the original sequence data;

FIG. 2 further elaborates that portion of the process schematized in FIG. 1 for predicting functional regions from genomic sequence;

FIG. 3 illustrates a Mondrian visual display;

FIG. 4 presents a Mondrian showing a hypothetical annotated genomic sequence;

FIG. 5 is a histogram showing the distribution of ORF length and PCR products as obtained, with ORF length shown in black and PCR product length shown in dotted lines;

FIG. 6 is a histogram showing the distribution, among exons predicted according to the methods described, of expression as measured using simultaneous two color hybridization to a genome-derived single exon microarray. The graph shows the number of sequence-verified products that were either not expressed ("0"), expressed in one or more but not all tested tissues ("1" - "9"), or expressed

in all tissues tested ("10");

FIG. 7 is a pictorial representation of the expression of verified sequences that showed expression with signal intensity greater than 3 in at least one tissue, with: FIG. 7A showing the expression as measured by microarray hybridization in each of the 10 measured tissues, and the expression as measured "bioinformatically" by query of EST, NR and SwissProt databases; with FIG. 7B showing the legend for display of physical expression (ratio) in FIG. 7A; and with FIG. 7C showing the legend for scoring EST hits as depicted in FIG. 7A;

FIG. 8 shows a comparison of normalized CY3 signal intensity for arrayed sequences that were identical to sequences in existing EST, NR and SwissProt databases or that were dissimilar (unknown), where black denotes the signal intensity for all sequence-verified products with a BLAST Expect ("E") value of greater than 1e-30 (1×10^{-30}) ("unknown") and a dotted line denotes sequence-verified spots with a BLAST expect ("E") value of less than 1e-30 (1×10^{-30}) ("known");

FIG. 9 presents a Mondrian of BAC AC008172 (bases 25,000 to 130,000), containing the carbamyl phosphate synthetase gene (AF154830.1); and

FIG. 10 is a Mondrian of BAC A049839.

Methods and Apparatus for Predicting, Confirming, Annotating, and Displaying Functional Regions From Genomic Sequence Data

FIG. 1 is a flow chart illustrating in broad outline a process for predicting functional regions from genomic sequence, confirming and characterizing the functional activity of such regions experimentally, and then associating and displaying the information so obtained

in meaningful and useful relationship to the original sequence data.

The initial input into process 10 of the present invention is drawn from one or more databases 100 containing genomic sequence data. Because genomic sequence is usually obtained from subgenomic fragments, the sequence data typically will be stored in a series of records corresponding to these subgenomic sequenced fragments. Some fragments will have been catenated to form larger contiguous sequences ("contigs"); others will not. A finite percentage of sequence data in the database will typically be erroneous, consisting *inter alia* of vector sequence, sequence created from aberrant cloning events, sequence of artificial polylinkers, and sequence that was erroneously read.

Each sequence record in database 100 will minimally contain as annotation a unique sequence identifier (accession number), and will typically be annotated further to identify the date of accession, species of origin, and depositor. Because database 100 can contain nongenomic sequence, each sequence will typically be annotated further to permit query for genomic sequence. Chromosomal origin, optionally with map location, can also be present. Data can be, and over time increasingly will be, further annotated with additional information, in part through use of the present invention, as described below. Annotation can be present within the data records, in information external to database 100 and linked to the records thereto, or through a combination of the two.

Databases useful as genomic sequence database 100 in the present invention include GenBank, and particularly include several divisions thereof, including the htgs(draft), NT (nucleotide, command line), and NR (nonredundant) divisions. GenBank is produced by the National Institutes of Health and is maintained by the

National Center for Biotechnology Information (NCBI). Databases of genomic sequence from species other than human, such as mouse, rat, Arabidopsis, *C. elegans*, *C. brigsii*, *Drosophila*, zebra fish, and other higher eukaryotic organisms will also prove useful as genomic sequence database 100.

Genomic sequence obtained by query of genomic sequence database 100 is then input into one or more processes 200 for identification of regions therein that are predicted to have a biological function as specified by the user. Such functions include, but are not limited to, encoding protein, regulating transcription, regulating message transport after transcription into mRNA, regulating message splicing after transcription into mRNA, of regulating message degradation after transcription into mRNA, and the like. Other functions include directing somatic recombination events, contributing to chromosomal stability or movement, contributing to allelic exclusion or X chromosome inactivation, and the like.

The particular genomic sequence to be input into process 200 will depend upon the function for which relevant sequence is to be identified as well as upon the approach chosen for such identification. Process step 200 can be iterated to identify different functions within a given genomic region. In such case, the input often will be different for the several iterations.

Sequences predicted to have the requisite function by process 200 are then input into process 300, where a subset of the input sequences suitable for experimental confirmation is identified. Experimental confirmation can involve physical and/or bioinformatic assay. Where the subsequent experimental assay is bioinformatic, rather than physical, there are fewer constraints on the sequences that can be tested, and in this latter case therefore process 300 can output the

entirety of the input sequence.

The subset of sequences output from process 300 is then used in process 400 for experimental verification and characterization of the function predicted in process 200, which experimental verification can, and often will, include both physical and bioinformatic assay.

Process 500 annotates the sequence data with the functional information obtained in the physical and/or bioinformatic assays of process 400. Such annotation can be done using any technique that usefully relates the functional information to the sequence, as, for example, by incorporating the functional data into the sequence data record itself, by linking records in a hierarchical or relational database, by linking to external databases, by a combination thereof, or by other means well known within the database arts. The data can even be submitted for incorporation into databases maintained by others, such as GenBank, which is maintained by NCBI.

As further noted in FIG. 1, additional annotation can be input into process 500 from external sources 600.

The annotated data is then displayed in process 800, either before, concomitantly with, or after optional storage 700 on nontransient media, such as magnetic disk, optical disc, magnetooptical disk, flash memory, or the like.

FIG. 1 shows that the experimental data output from process 400 can be used in each preceding step of process 10: e.g., facilitating identification of functional sequences in process 200, facilitating identification of an experimentally suitable subset thereof in process 300, and facilitating creation of physical and/or informational substrates for, and performance of subsequent assay, of functional sequences in process 400.

Information from each step can be passed directly to the succeeding process, or stored in permanent or

interim form prior to passage to the succeeding process. Often, data will be stored after each, or at least a plurality, of such process steps. Any or all process steps can be automated.

FIG. 2 further elaborates the prediction of functional sequence within genomic sequence according to process 200.

Genomic sequence database 100 is first queried 20 for genomic sequence.

The sequence required to be returned by query 20 will depend, in the first instance, upon the function to be identified.

For example, genomic sequences that function to encode protein can be identified *inter alia* using gene prediction approaches, comparative sequence analysis approaches, or combinations of the two. In gene prediction analysis, sequence from one genome is input into process 200 where at least one, preferably a plurality, of algorithmic methods are applied to identify putative coding regions. In comparative sequence analysis, by contrast, corresponding, *e.g.*, syntenic, sequence from a plurality of sources, typically a plurality of species, is input into process 200, where at least one, possibly a plurality, of algorithmic methods are applied to compare the sequences and identify regions of least variability.

The exact content of query 20 will also depend upon the database queried. For example, if the database contains both genomic and nongenomic sequence, perhaps derived from multiple species, and the function to be determined is protein coding regions in human genomic sequence, the query will accordingly require that the sequence returned be genomic and derived from humans.

Query 20 can also incorporate criteria that compel return of sequence that meets operative requirements of the subsequent analytical method. Alternatively, or in

addition, such operative criteria can be enforced in subsequent preprocess step 24.

For example, if the function sought to be identified is protein coding, query 20 can incorporate criteria that return from genomic sequence database 100 only those sequences present within contigs sufficiently long as to have obviated substantial fragmentation of any given exon among a plurality of separate sequence fragments.

Such criteria can, for example, consist of a required minimal individual genomic sequence fragment length, such as 10 kb, more typically 20 kb, 30 kb, 40kb, and preferably 50 kb or more, as well as an optional further or alternative requirement that sequence from any given clone, such as a bacterial artificial chromosome ("BAC"), be presented in no more than a finite maximal number of fragments, such as no more than 20 separate pieces, more typically no more than 15 fragments, even more typically no more than about 10 - 12 fragments.

Results using the present invention have shown that genomic sequence from bacterial artificial chromosomes (BACs) is sufficient for gene prediction analysis according to the present invention if the sequence is at least 50 kb in length, and if additionally the sequence from any given BAC is presented in fewer than 15, and preferably fewer than 10, fragments. Accordingly, query 20 can incorporate a requirement that data accessioned from BAC sequencing be in fewer than 15, preferably fewer than 10, fragments.

An additional criterion that can be incorporated into the query can be the date, or range of dates, of sequence accession. Although the process has been described above as if genomic sequence database 100 were static, it is of course understood that the genomic sequence databases need not be static, and indeed are typically updated on a frequent, even hourly, basis. Thus,

as further described in Examples 1 and 2, *infra*, it is possible to query the database for newly added sequence, either newly added after an absolute date, or newly added relative to a prior analysis performed using the methods and apparatus of the present invention. In this way, the process herein described can incorporate a dynamic, temporal component.

One utility of such temporal limitation is to identify, from newly accessioned genomic sequence, the presence of novel genes, particularly those not previously identified by EST sequencing (or other sequencing efforts that are similarly based upon gene expression). As further described in Example 1, such an approach has shown that newly accessioned human genomic sequence, when analyzed for sequences that function to encode protein, readily identifies genes that are novel over those in existing EST and other expression databases. This makes the methods of the present invention extremely powerful gene discovery tools. And as would be appreciated, such gene discovery can be performed using genomic sequence from species other than human.

If query 20 incorporates multiple criteria, such as above-described, the multiple criteria can be performed as a series of separate queries or as a single query, depending in part upon the query language, the complexity of the query, and other considerations well known in the database arts.

If query 20 returns no genomic sequence meeting the query criteria, the negative result can be reported by process 22, and process 200 (and indeed, entire process 10) ended 23, as shown. Alternatively, or in addition to report and termination of the initial inquiry, a new query 20 can be generated that takes into account the initial negative result.

When query 20 returns sequence meeting the query

criteria, the returned sequence is then passed to optional preprocessing 24, suitable and specific for the desired analytical approach and the particular analytical methods thereof to be used in process 25.

Preprocessing 24 can include processes suitable for many approaches and methods thereof, as well as processes specifically suited for the intended subsequent analysis.

Preprocessing 24 suitable for most approaches and methods will include elimination of sequence irrelevant to, or that would interfere with, the subsequent analysis. Such sequence includes repetitive sequence, such as Alu repeats and LINE elements, vector sequence, artificial sequence, such as artificial polylinkers, and the like. Such removal can readily be performed by identification and subsequent masking of the undesired sequence.

Identification can be effected by comparing the genomic sequence returned by query 20 with public or private databases containing known repetitive sequence, vector sequence, artificial sequence, and other artifactual sequence. Such comparison can readily be done using programs well known in the art, such as CROSS_MATCH, or by proprietary sequence comparison programs the engineering of which is well within the skill in the art.

Alternatively, or in addition, undesirable, including artifactual, sequence can be identified algorithmically without comparison to external databases and thereafter removed. For example, synthetic polylinker sequence can be identified by an algorithm that identifies a significantly higher than average density of known restriction sites. As another example, vector sequence can be identified by algorithms that identify nucleotide or codon usage at variance with that of the bulk of the genomic sequence.

Once identified, undesired sequence can be

removed. Removal can usefully be done by masking the undesired sequence as, for example, by converting the specific nucleotide references to one that is unrecognized by the subsequent bioinformatic algorithms, such as "X". Alternatively, but at present less preferred, the undesired sequence can be excised from the returned genomic sequence, leaving gaps.

Preprocessing 24 can further include selection from among duplicative sequences of that one sequence of highest quality. Higher quality can be measured as a lower percentage of, fewest number of, or least densely clustered occurrence of ambiguous nucleotides, defined as those nucleotides that are identified in the genomic sequence using symbols indicating ambiguity. Higher quality can also or alternatively be valued by presence in the longest contig.

Preprocessing 24 can, and often will, also include formatting of the data as specifically appropriate for passage to the analytical algorithms of process 25. Such formatting can and typically will include, *inter alia*, addition of a unique sequence identifier, either derived from the original accession number in genomic sequence database 100, or newly applied, and can further include additional annotation. Formatting can include conversion from one to another sequence listing standard, such as conversion to or from FASTA or the like, depending upon the input expected by the subsequent process.

Preprocessing, which can be optional depending upon the function desired to be identified and the informational requirements of the methods for effecting such identification, is followed by sequence processing 25, where sequences with the desired function are identified within the genomic sequence.

As mentioned above, such functions can include, but are not limited to, encoding protein, regulating

transcription, regulating message transport after transcription into mRNA, regulating message splicing after transcription, of regulating message degradation, and the like. Other functions include directing somatic recombination events, contributing to chromosomal stability or movement, contributing to allelic exclusion or X chromosome inactivation, or the like.

The methods of the present invention are particularly useful for gene discovery, that is, for identifying, from genomic sequence, regions that function to encode genes, and in a particularly useful embodiment, for identifying regions that function to encode genes not hitherto identified by expression-based or directed cloning and sequencing. In conjunction with verification using the novel single exon microarrays of the present invention, as further described below, the methods herein described become powerful gene discovery tools.

Accordingly, in a preferred embodiment of the present invention, process 25 is used to identify putative coding regions. Two preferred approaches in process 25 for identifying sequence that encodes putative genes are gene prediction and comparative sequence analysis.

Gene prediction can be performed using any of a number of algorithmic methods, embodied in one or more software programs, that identify open reading frames (ORFs) using a variety of heuristics, such as GRAIL, DICTION, and GENEFINDER. Comparative sequence analysis similarly can be performed using any of a variety of known programs that identify regions with lower sequence variability.

As further described in Example 1, below, gene finding software programs yield a range of results. For the newly accessioned human genomic sequence input in Example 1, for example, GRAIL identified the greatest percentage of genomic sequence as putative coding region, 2% of the data analyzed; GENEFINDER was second, calling 1%;

and DICTION yielded the least putative coding region, with 0.8% of genomic sequence called as coding region.

Increased reliability can be obtained when consensus is required among several such methods. Although discussed herein particularly with respect to exon calling, consensus among methods will in general increase reliability of predicting other functions as well.

Thus, as indicated by query 26, sequence processing 25, optionally with preprocessing 24, can be repeated with a different method, with consensus among such iterations determined and reported in process 27.

Process 27 compares the several outputs for a given input genomic sequence and identifies consensus among the separately reported results. The consensus itself, as well as the sequence meeting that consensus, is then stored in process 29a, displayed in process 29b, and/or output to process 300 for subsequent identification of a subset thereof suitable for assay.

Multiple levels of consensus can be calculated and reported by process 27. For example, as further described in Example 1, *infra*, process 27 can report consensus as between all specific pairs of methods of gene prediction, as consensus among any one or more of the pairs of methods of gene prediction, or as among all of the gene prediction algorithms used. Thus, in Example 1, process 27 reported that GRAIL and GENEFINDER programs agreed on 0.7% of genomic sequence, that GRAIL and DICTION agreed on 0.5% of genomic sequence, and that the three programs together agreed on 0.25% of the data analyzed. Put another way, 0.25% of the genomic sequence was identified by all three of the programs as containing putative coding region.

Furthermore, consensus can be required among different approaches to identifying a chosen function.

For example, if the function desired to be identified is coding of protein sequence, and a first used

approach to exon calling is gene prediction, the process can be repeated on the same input sequence, or subset thereof, with another approach, such as comparative sequence analysis. In such a case, where comparative sequence analysis follows gene prediction, the comparison can be performed not only on genomic nucleic acid sequence, but additionally or alternatively can be performed on the predicted amino acid sequence translated from the ORFs prior identified by the gene prediction approach.

Although shown as an iterative process, the multiple analyses required to achieve consensus can be done in series, in parallel, or some combination thereof.

Predicted functional sequence, optionally representing a consensus among a plurality of methods and approaches for determination thereof, is passed to process 300 for identification of a subset thereof for functional assay.

In the preferred embodiment of the methods of the present invention, wherein the function sought to be identified is protein coding, process 300 is used to identify a subset thereof suitable for experimental verification by physical and/or bioinformatic approaches.

For example, putative ORFs identified in process 200 can be classified, or binned, bioinformatically into putative genes. This binning can be based *inter alia* upon consideration of the average number of exons/gene in the species chosen for analysis, upon density of exons that have been called on the genomic sequence, and other empirical rules. Thereafter, one or more among the gene-specific ORFs can be chosen for subsequent use in gene expression assay.

Where such subsequent gene expression assay uses amplified nucleic acid, considerations such as desired amplicon length, primer synthesis requirements, putative exon length, sequence GC content, existence of possible

secondary structure, and the like can be used to identify and select those ORFs that appear most likely successfully to amplify. Where subsequent gene expression assay relies upon nucleic acid hybridization, whether or not using amplified product, further considerations involving hybridization stringency can be applied to identify that subset of sequences that will most readily permit sequence-specific discrimination at a chosen hybridization and wash stringency. One particular such consideration is avoidance of putative exons that span repetitive sequence; such sequence can hybridize spuriously to nonspecific message, reducing specific signal in the hybridization.

For bioinformatic assay, there are fewer constraints on the sequences that can be tested experimentally, and in this latter case therefore process 300 can output the entirety of the input sequence.

The subset of sequences identified by process 300 as suitable for use in assay is then used in process 400 to create the physical and/or informational substrate for experimental verification of the predictions made in process 200, and thereafter to assay those substrates.

As mentioned, the methods of the present invention are particularly useful for identifying potential coding regions within genomic sequence. In a preferred embodiment of process 400, therefore, the expression of the sequences predicted to encode protein is verified. The combination of the predictive and experimental methods provides a powerful gene discovery engine.

Thus, in another aspect, the present invention provides methods and apparatus for verifying the expression of putative genes identified within genomic sequence. In particular, the invention provides a novel method of verifying gene expression in which expression of predicted ORFs is measured and confirmed using a novel type of nucleic acid microarray, the genome-derived single exon

nucleic acid microarrays of the present invention.

Putative ORFs as predicted by a consensus of gene calling, particularly gene prediction, algorithms in process 200, and as further identified as suitable by process 300, are amplified from genomic DNA using the polymerase chain reaction (PCR). Although PCR is conveniently used, other amplification approaches can also be used.

Amplification schemes can be designed to capture the entirety of each predicted ORF in an amplicon with minimal additional (that is, intronic or intergenic) sequence. Because ORFs predicted from human genomic sequence using the methods of the present invention differ in length, such an approach results in amplicons of varying length.

However, most predicted ORFs are shorter than 500 bp in length, and although amplicons of at least about 100 or 200 base pairs can be immobilized as probes on nucleic acid microarrays, early experimental results using the methods of the present invention have suggested that longer amplicons, at least about 400 or 500 base pairs, are more effective. Furthermore, certain advantages derive from application to the microarray of amplicons of defined size.

Therefore, amplification schemes can alternatively, and preferably, be designed to amplify regions of defined size, preferably at least about 300, 400 or 500 bp, centered about each predicted ORF. Such an approach results in a population of amplicons of limited size diversity, but that typically contain intronic and/or intergenic nucleic acid in addition to putative ORF.

Conversely, somewhat fewer than 10% of ORFs predicted from human genomic sequence according to the methods of the present invention exceed 500 bp in length. Portions of such extended ORFs, preferably at least about 300,400 or 500 bp in length, can be amplified. However, it

has been discovered that the percentage success at amplifying pieces of such ORFs is low, and that such putative exons are more effectively amplified when larger fragments, at least about 1000 or 1500 bp, and even as large as 2000 bp are amplified.

The putative ORFs selected in process 300 are thus input into one or more primer design programs, such as PRIMER3 (available online for use at http://www-genome.wi.mit.edu/cgi-bin/primer/), with a goal of amplifying at least about 500 base pairs of genomic sequence centered within or about ORFs predicted to be no more than about 500 bp, or at least about 1000 - 1500 bp of genomic sequence for ORFs predicted to exceed 500 bp in length, and the primers synthesized by standard techniques. Primers with the requisite sequences can be purchased commercially or synthesized by standard techniques.

Conveniently, a first predetermined sequence can be added commonly to the ORF-specific 5' primer and a second, typically different, predetermined sequence commonly added to each 3' ORF-unique primer. This serves to immortalize the amplicon, that is, serves to permit further amplification of any amplicon using a single set of primers complementary respectively to the common 5' and common 3' sequence elements. The presence of these "universal" priming sequences further facilitates later sequence verification, providing a sequence common to all amplicons at which to prime sequencing reactions. The common 5' and 3' sequences further serve to add a cloning site should any of the ORFs warrant further study.

Such predetermined sequence is usefully at least about 10, 12 or 15 nt in length, and usually does not exceed about 25 nt in length. The "universal" priming sequences used in the examples presented *infra* were each 16 nt long.

The genomic DNA to be used as substrate for

amplification will come from the eukaryotic species from which the genomic sequence data had originally been obtained, or a closely related species, and can conveniently be prepared by well known techniques from somatic or germline tissue or cultured cells of the organism. See, *e.g.*, Short Protocols in Molecular Biology : A Compendium of Methods from Current Protocols in Molecular Biology, Ausubel et al. (eds.), 4th edition (April 1999), John Wiley & Sons (ISBN: 047132938X) and Maniatis *et al.*, Molecular Cloning : A Laboratory Manual, 2nd edition (December 1989), Cold Spring Harbor Laboratory Press (ISBN: 0879693096). Many such prepared genomic DNAs are available commercially, with the human genomic DNAs additionally having certification of donor informed consent.

Although the intronic and intergenic material flanking putative coding regions in the amplicons could potentially interfere with hybridizations during microarray experiments, we have found, surprisingly, that differential expression ratios are not significantly affected. Rather, the predominant effect of exon size is to alter the absolute signal intensity, rather than its ratio. Equally surprising, the art had suggested that single exon probes would not provide sufficient signal intensity for high stringency hybridization analyses; we find that such probes not only provide adequate signal, but have substantial advantages, as herein described.

After partial purification, as by size exclusion spin column, with or without confirmation as to amplicon quality as by gel electrophoresis, each amplicon (single exon probe) is disposed in an array upon a support substrate.

Methods for creating microarrays by deposition and fixation of nucleic acids onto support substrates are well known in the art (Reviewed by Schena et al., see

above).

Typically, the support substrate will be glass, although other materials, such as amorphous or crystalline silicon or plastics. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof, can also be used. Typically, the support will be rectangular, although other shapes, particularly circular disks and even spheres, present certain advantages. Particularly advantageous alternatives to glass slides as support substrates for array of nucleic acids are optical discs, as described in WO 98/12559.

The amplified nucleic acids can be attached covalently to a surface of the support substrate or, more typically, applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof.

Robotic spotting devices useful for arraying nucleic acids on support substrates can be constructed using public domain specifications (The MGuide, version 2.0, http://cmgm.stanford.edu/pbrown/mguide/index.html), or can conveniently be purchased from commercial sources (MicroArray GenII Spotter and MicroArray GenIII Spotter, Molecular Dynamics, Inc., Sunnyvale, CA). Spotting can also be effected by printing methods, including those using ink jet technology.

As is well known in the art, microarrays typically also contain immobilized control nucleic acids. For controls useful in providing measurements of background signal for the genome-derived single exon microarrays of the present invention, a plurality of *E. coli* genes can

readily be used. As further described in Example 1, 16 or 32 *E. coli* genes suffice to provide a robust measure of background noise in such microarrays.

As is well known in the art, the amplified product disposed in arrays on a support substrate to create a nucleic acid microarray can consist entirely of natural nucleotides linked by phosphodiester bonds, or alternatively can include either nonnative nucleotides, alternative internucleotide linkages, or both, so long as complementary binding can be obtained in the hybridization. If enzymatic amplification is used to produce the immobilized probes, the amplifying enzyme will impose certain further constraints upon the types of nucleic acid analogs that can be generated.

Although particularly described herein as using high density microarrays constructed on planar substrates, the methods of the present invention for confirming the expression of ORFs predicted from genomic sequence can use any of the known types of microarrays, as herein defined, including lower density planar arrays, and microarrays on nonplanar, nonunitary, distributed substrates.

For example, gene expression can be confirmed using hybridization to lower density arrays, such as those constructed on membranes, such as nitrocellulose, nylon, and positively-charged derivatized nylon membranes. Further, gene expression can also be confirmed using nonplanar, bead-based microarrays such as are described in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000); U.S. Patent No. 6,057,107; and U.S. Patent No. 5,736,330. In theory, a packed collection of such beads provides in aggregate a higher density of nucleic acid probe than can be achieved with spotting or lithography techniques on a single planar substrate.

Planar microarrays on solid substrates, however, provide certain useful advantages, including high

throughput and compatibility with existing readers. For example, each standard microscope slide can include at least 1000, typically at least 2000, preferably 5000 and upto 10,000 - 50,000 or more nucleic acid probes of discrete sequence. The number of sequences deposited will depend on their required application.

Each putative gene can be represented in the array by a single predicted ORF. Alternatively, genes can be represented by more than one predicted ORF. For purposes of measuring differential splicing, more than one predicted ORF will be provided for a putative gene. And as is well known in the art, each probe of defined sequence, representing a single predicted ORF, can be deposited in a plurality of locations on a single microarray to provide redundancy of signal.

The genome-derived single exon microarrays described above differ in several fundamental and advantageous ways from microarrays presently used in the gene expression art, including (1) those created by deposition of mRNA-derived nucleic acids, (2) those created by *in situ* synthesis of oligonucleotide probes, and (3) those constructed from yeast genomic DNA.

Most nucleic acid microarrays that are in use for study of eukaryotic gene expression have as immobilized probes nucleic acids that are derived – either directly or indirectly – from expressed message. As discussed above, it is common, for example, for such microarrays to be derived from cDNA/EST libraries, either from those previously described in the literature, see Lennon *et al.*, or from the *de novo* construction of "problem specific" libraries targeted at a particular biological question, R.S. Thomas *et al.*, *Cancer Res.* (in press). Such microarrays are herein collectively denominated "EST microarrays".

Such EST microarrays by definition can measure

expression only of those genes found in EST libraries, shown herein to represent only a fraction of expressed genes. Furthermore, such libraries – and thus microarrays based thereupon – are biased by the tissue or cell type of message origin, by the expression levels of the respective genes within the tissues, and by the ability of the message successfully to have been reverse-transcribed and cloned.

Thus, as further discussed in Example 1, the methods of the present invention enable sequences that do not appear in EST or other expression databases to be determined - subsequently arrayed for expression measurements could not, therefore, have been represented as probes on an EST microarray. And as further demonstrated in the examples, *infra*, the remaining population of genes identified from genomic sequence by the methods of the present invention – that is, the one third of sequences that had previously been accessioned in EST or other expression databases – are biased toward genes with higher expression levels.

Representation of a message in an EST and/or cDNA library depends upon the successful reverse transcription, optionally but typically with subsequent successful cloning, of the message. This introduces substantial bias into the population of probes available for arraying in EST microarrays.

In contrast, neither reverse transcription nor cloning is required to produce the probes arrayed on the genome-derived single exon microarrays of the present invention. And although the ultimate deposition of a probe on the genome-derived single exon microarray of the present invention depends upon a successful amplification from genomic material, *a priori* knowledge of the sequence of the desired amplicon affords greater opportunity to recover any given probe sequence recalcitrant to amplification than is afforded by the requirement for successful reverse

transcription and cloning of unknown message in EST approaches.

Thus, the genome-derived single exon microarrays of the present invention present a far greater diversity of probes for measuring gene expression, with far less bias, than do EST microarrays presently used in the art.

As a further consequence of their ultimate origin from expressed message, the probes in EST microarrays often contain poly-A (or complementary poly-T) stretches derived from the poly-A tail of mature mRNA. These homopolymeric stretches contribute to cross-hybridization, that is, to a spurious signal occasioned by hybridization to the homopolymeric tail of a labeled cDNA that lacks sequence homology to the gene-specific portion of the probe.

In contrast, the probes arrayed in the genome-derived single exon microarrays of the present invention lack homopolymeric stretches derived from message polyadenylation, and thus can provide more specific signal. Typically, at least about 50, 60 or 75% of the probes on the genome-derived single exon microarrays of the present invention lack homopolymeric regions consisting of A or T, where a homopolymeric region is defined for purposes herein as stretches of 25 or more, typically 30 or more, identical nucleotides.

A further distinction, which also affects the specificity of hybridization, is occasioned by the typical derivation of EST microarray probes from cloned material. Because much of the probe material disposed as probes on EST microarrays is excised or amplified from plasmid, phage, or phagemid vectors, EST microarrays typically include a fair amount of vector sequence, more so when the probes are amplified, rather than excised, from the vector.

In contrast, the vast majority of probes in the genome-derived single exon microarrays of the present invention contain no prokaryotic or bacteriophage vector

sequence, having been amplified directly or indirectly from genomic DNA. Typically, therefore, at least about 50, 60, 70 or 80% or more of individual exon-including probes disposed on a genome-derived single exon microarray of the present invention lack vector sequence, and particularly lack sequences drawn from plasmids and bacteriophage. Preferably, at least about 85, 90 or more than 90% of exon-including probes in the genome-derived single exon microarray of the present invention lack vector sequence. With attention to removal of vector sequences through preprocessing 24, percentages of vector-free exon-including probes can be as high as 95 - 99%. The substantial absence of vector sequence from the genome-derived single exon microarrays of the present invention results in greater specificity during hybridization, since spurious cross-hybridization to a probe vector sequence is reduced.

As a further consequence of excision or amplification of probes from vectors in construction of EST microarrays, the probes arrayed thereon often contain artificial sequence, derived from vector polylinker multiple cloning sites, at both 5' and 3' ends. The probes disposed upon the genome-derived single exon microarrays need have no such artificial sequence appended thereto.

As mentioned above, however, the ORF-specific primers used to amplify putative ORFs can include artificial sequences, typically 5' to the ORF-specific primer sequence, useful for "universal" (that is, independent of ORF sequence) priming of subsequent amplification or sequencing reactions. When such "universal" 5' and/or 3' priming sequences are appended to the amplification primers, the probes disposed upon the genome-derived single exon microarray will include artificial sequence similar to that found in EST microarrays. However, the genome-derived single exon microarray of the present invention can be made without

such sequences, and if so constructed, presents an even smaller amount of nonspecific sequence that would contribute to nonspecific hybridization.

Yet another consequence of typical use of cloned material as probes in EST microarrays is that such microarrays contain probes that result from cloning artifacts, such as chimeric molecules containing coding region of two separate genes. Derived from genomic material, typically not thereafter cloned, the probes of the genome-derived single exon microarrays of the present invention lack such cloning artifacts, and thus provide greater specificity of signal in gene expression measurements.

A further consequence of the cloned origin of probes on many EST microarrays is that the individual probes often have disparate sizes, which can cause the optimal hybridization stringency to vary among probes on a single microarray. In contrast, as discussed above, the probes arrayed on the genome-derived single exon microarrays of the present invention can readily be designed to have a narrow distribution in sizes, with the range of probe sizes no greater than about 10% of the average size, typically no greater than about 5% of the average probe size.

Because of their origin from fully- or partially-spliced message, probes disposed upon EST arrays will often include multiple exons. The percentage of such exon-spanning probes in an EST microarray can be calculated, on average, based upon the predicted number of exons/gene for the given species and the average length of the immobilized probes. For human genes, the near-complete sequence of human chromosome 22, Dunham *et al.*, *Nature* 402(6761):489-95 (1999), predicts that human genes average 5.5 exons/gene. Even with probes of 200 - 500 bp, the vast majority of human EST microarray probes include more than one exon.

In contrast, by virtue of their origin from algorithmically identified ORFs in genomic sequence, the probes in the genome-derived single exon microarrays of the present invention can consist of individual exons. Thus, in contrast to EST microarrays, at least about 50, 60, 70, 75, 80, 85, 95 or 99% of probes deposited in the genome-derived microarray of the present invention consist of, or include, no more than one predicted ORF.

This provides the ability, not readily achieved using EST microarrays, to use the genome-derived single exon microarrays of the present invention to measure tissue-specific expression of individual exons, which in turn allows differential splicing events to be detected and characterized, and in particular, allows the correlation of differential splicing to tissue-specific expression patterns.

Furthermore, the exons that are represented in EST microarrays are often biased toward the 3' or 5' end of their respective genes, since sequencing strategies used for EST identification are so biased. In contrast, no such 3' or 5' bias necessarily inheres in the selection of exons for disposition on the genome-derived single exon microarrays of the present invention.

Conversely, the probes provided on the genome-derived single exon microarrays of the present invention typically, but need not necessarily, include intronic and/or intergenic sequence that is absent from EST microarrays, which are derived from mature mRNA. Typically, at least about 50, 60, 70, 80 or 90% of the exon-including probes on the genome-derived single exon microarrays of the present invention include sequence drawn from noncoding regions. As discussed above, the additional presence of noncoding region does not significantly interfere with measurement of gene expression, and provides the additional opportunity to assay prespliced RNA, and

thus measure such phenomena such as nuclear export control.

The genome-derived single exon microarrays of the present invention are also quite different from *in situ* synthesis microarrays, where probe size is severely constrained by inadequacies in the photolithographic synthesis process.

Typically, probes arrayed on *in situ* synthesis microarrays are limited to a maximum of about 25 bp. As a well known consequence, hybridization to such chips must be performed at low stringency. In order, therefore, to achieve unambiguous sequence-specific hybridization results, the *in situ* synthesis microarray requires substantial redundancy, with concomitant programmed arraying for each probe of probe analogues with altered (*i.e.*, mismatched) sequence.

In contrast, the longer probe length of the genome-derived single exon microarrays of the present invention allows much higher stringency hybridization and wash. Typically, therefore, exon-including probes on the genome-derived single exon microarrays of the present invention average at least about 100, 200, 300, 400 or 500 bp in length. By obviating the need for substantial probe redundancy, this approach permits a higher density of probes for discrete exons or genes to be arrayed on the microarrays of the present invention than can be achieved for *in situ* synthesis microarrays.

A further distinction is that the probes in *in situ* synthesis microarrays typically are covalently linked to the substrate surface. In contrast, the probes disposed on the genome-derived microarray of the present invention typically are, but need not necessarily be, bound noncovalently to the substrate.

Furthermore, the short probe size on *in situ* microarrays causes large percentage differences in the melting temperature of probes hybridized to their

complementary target sequence, and thus causes large percentage differences in the theoretically optimum stringency across the array as a whole.

In contrast, the larger probe size in the microarrays of the present invention create lower percentage differences in melting temperature across the range of arrayed probes.

A further significant advantage of the microarrays of the present invention over *in situ* synthesized arrays is that the quality of each individual probe can be confirmed before deposition. In contrast, the quality of probes cannot be assessed on a probe-by-probe basis for the *in situ* synthesized microarrays presently being used.

The genome-derived single exon microarrays of the present invention are also distinguished over, and present substantial benefits over, the genome-derived microarrays from lower eukaryotes such as yeast. Lashkari *et al.*, *Proc. Natl. Acad. Sci. USA* 94:13057-13062 (1997).

Only about 220 - 250 of the 6100 or so nuclear genes in *Saccharomyces cerevisiae* – that is, only about 4 - 5% – have standard, spliceosomal, introns, Lopez *et al.*, *Nucl. Acids Res.* 28:85-86 (2000); Spingola *et al.*, *RNA* 5(2):221-34 (1999). Furthermore, the entire yeast genome has already been sequenced. These two facts permit the ready amplification and disposition of single-ORF amplicons on such microarray without the requirement for antecedent use of gene prediction and/or comparative sequence analyses.

Thus, a significant aspect of the present invention is the ability to identify and to confirm expression of predicted coding regions in genomic sequence drawn from eukaryotic organisms that have a higher percentage of genes having introns than do yeast such as *Saccharomyces cerevisiae*, particularly in genomic sequence

drawn from eukaryotes in which at least about 10, 20 or 50% of protein-encoding genes have introns. In preferred embodiments, the methods and apparatus of the present invention are used to identify and confirm expression of novel genes from genomic sequence of eukaryotes in which the average number of introns per gene is at least about one, two or three or more.

After the physical substrate is prepared, experimental verification of predicted function is performed.

In a preferred embodiment of the present invention, where the function sought to be identified in genomic sequence is protein coding, experimental verification is performed by measuring expression of the putative ORFs, typically through nucleic acid hybridization experiments, and in particularly preferred embodiments, through hybridization to genome-derived single exon microarrays prepared as above- described.

Expression is conveniently measured and expressed for each probe in the microarray as a ratio of the expression measured concurrently in a plurality of mRNA sources, according to techniques well known in the microarray art, *Reviewed in* Schena et al., and as further described in Example 2, below. The mRNA source for the reference against which specific expression is measured can be drawn from a homogeneous mRNA source, such as a single cultured cell-type, or alternatively can be heterogeneous, as from a pool of mRNA derived from multiple tissues and/or cell types, as further described in Example 2, *infra*.

mRNA can be prepared by standard techniques, see Ausubel et al. and Maniatis *et al.*, or purchased commercially. The mRNA is then typically reverse-transcribed in the presence of labeled nucleotides: the index source (that in which expression is desired to be measured) is reverse transcribed in the presence of

nucleotides labeled with a first label, typically a fluorophore (fluorochrome; fluor; fluorescent dye); the reference source is reverse transcribed in the presence of a second label, typically a fluorophore, typically fluorometrically-distinguishable from the first label. As further described in Example 2, *infra*, Cy3 and Cy5 dyes prove particularly useful in these methods. After partial purification of the index and reference targets, hybridization to the probe array is conducted according to standard techniques, typically under a coverslip.

After wash, microarrays are conveniently scanned using a commercial microarray scanning device, such as a Gen3 Scanner (Molecular Dynamics, Sunnyvale, CA). Data on expression is then passed, with or without interim storage, to process 500, where the results for each probe are related to the original sequence.

Often, hybridization of target material to the genome-derived single exon microarray will identify certain of the probes thereon as of particular interest. Thus, it is often desirable that the user be able readily to obtain sufficient quantities of an individual probe, either for subsequent arrayed deposition upon an additional support substrate, often as part of a microarray having a plurality of probes so identified, or alternatively or additionally as a solitary solid-phase or solution-phase probe, for further use.

Thus, in another aspect, the present invention provides compositions and kits for the ready production of nucleic acids identical in sequence to, or substantially identical in sequence to, probes on the genome-derived single exon microarrays of the present invention.

In this aspect, a small quantity of each probe is disposed, typically without attachment to substrate, in a spatially-addressable ordered set, typically one per well of a microtiter dish. Although a 96 well microtiter plate

can be used, greater efficiency is obtained using higher density arrays, such as are provided by microtiter plates having 384, 864, 1536, 3456, 6144, or 9600 wells, and although microtiter plates having physical depressions (wells) are conveniently used, any device that permits addressable withdrawal of reagent from fluidly-noncommunicating areas can be used.

In this aspect of the invention, therefore, a fluidly noncommunicating addressable ordered set of individual probes, corresponding to those on a genome-derived single exon microarray, is provided, with each probe in sufficient quantity to permit amplification, such as by PCR. As earlier mentioned, the ORF-specific 5' primers used for genomic amplification can have a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification can have a second, different, common sequence added thereto, thus permitting, in this preferred embodiment, the use of a single set of 5' and 3' primers to amplify any one of the probes from the amplifiable ordered set.

Each discrete amplifiable probe can also be packaged with amplification primers, solutes, buffers, *etc.*, and can be provided in dry (*e.g.*, lyophilized) form or wet, in the latter case typically with addition of agents that retard evaporation.

In another aspect of the present invention, a genome-derived single-exon microarray is packaged together with such an ordered set of amplifiable probes corresponding to the probes, or one or more subsets of probes, thereon. In alternative embodiments, the ordered set of amplifiable probes is packaged separately from the genome-derived single exon microarray.

In some embodiments, the microarray and/or ordered probe set are further packaged with recordable media that provide probe identification and addressing

information, and that can additionally contain annotation information, such as gene expression data. Such recordable media can be packaged with the microarray, with the ordered probe set, or with both.

If the microarray is constructed on a substrate that incorporates recordable media, such as is described in international patent application no. WO 98/12559, then separate packaging of the genome-derived single exon microarray and the bioinformatic information is not required.

The amount of amplifiable probe material should be sufficient to permit at least one amplification sufficient for subsequent hybridization assay.

Although the use of high density genome-derived microarrays on solid planar substrates is presently a preferred approach for the physical confirmation and characterization of the expression of sequences predicted to encode protein, other types of microarrays (as herein defined) can also be used.

Furthermore, as earlier mentioned, experimental verification of the function predicted from genomic sequence in process 200 can be bioinformatic, rather than, or additional to, physical verification.

For example, where the function desired to be identified is protein coding, the predicted ORFs can be compared bioinformatically to sequences known or suspected of being expressed.

Thus, the sequences output from process 300 (or process 200), can be used to query expression databases, such as EST databases, SNP ("single nucleotide polymorphism") databases, known cDNA and mRNA sequences, SAGE ("serial analysis of gene expression") databases, and more generalized sequence databases that allow query for expressed sequences. Such query can be done by any sequence query algorithm, such as BLAST ("basic local

alignment search tool"). The results of such query - including information on identical sequences and information on nonidentical sequences that have diffuse or focal regions of sequence homology to the query sequence - can then be passed directly to process 500, or used to inform analyses subsequently undertaken in process 200, process 300, or process 400.

Experimental data, whether obtained by physical or bioinformatic assay in process 400, is passed to process 500 where it is usefully related to the sequence data itself, a process colloquially termed "annotation". Such annotation can be done using any technique that usefully relates the functional information to the sequence, as, for example, by incorporating the functional data into the record itself, by linking records in a hierarchical or relational database, by linking to external databases, or by a combination thereof. Such database techniques are well within the skill in the art.

The annotated sequence data can be stored locally, uploaded to genomic sequence database 100, and/or displayed 800.

The methods and apparatus of the present invention rapidly produce functional information from genomic sequence. Coupled with the escalating pace at which sequence now accumulates, the rapid pace of sequence annotation produces a need for methods of displaying the information in meaningful ways.

FIG. 3 shows visual display 80 presenting a single genomic sequence annotated according to the present invention. Because of its nominal resemblance to artistic works of Piet Mondrian, visual display 80 is alternatively described herein as a "Mondrian".

Each of the visual elements of display 80 is aligned with respect to the genomic sequence being annotated (hereinafter, the "annotated sequence"). Given

the number of nucleotides typically represented in an annotated sequence, representation of individual nucleotides would rarely be readable in hard copy output of display 80. Typically, therefore, the annotated sequence is schematized as rectangle 89, extending from the left border of display 80 to its right border. By convention herein, the left border of rectangle 89 represents the first nucleotide of the sequence and the right border of rectangle 89 represents the last nucleotide of the sequence.

As further discussed below, however, the Mondrian visual display of annotated sequence can serve as a convenient graphical user interface for computerized representation, analysis, and query of information stored electronically. For such use, the individual nucleotides can conveniently be linked to the X axis coordinate of rectangle 89. This permits the annotated sequence at any point within rectangle 89 readily to be viewed, either automatically – for example, by time-delayed appearance of a small overlaid window upon movement of a cursor or other pointer over rectangle 89 – or through user intervention, as by clicking a mouse or other pointing device at a point in rectangle 89.

Visual display 80 is generated after user specification of the genomic sequence to be displayed. Such specification can consist of or include an accession number for a single clone (*e.g.*, a single BAC accessioned into GenBank), wherein the starting and stopping nucleotides are thus absolutely identified, or alternatively can consist of or include an anchor or fulcrum point about which a chosen range of sequence is anchored, thus providing relative endpoints for the sequence to be displayed. For example, the user can anchor such a range about a given chromosomal map location, gene name, or even a sequence returned by query for similarity

or identity to an input query sequence. When visual display 80 is used as a graphical user interface to computerized data, additional control over the first and last displayed nucleotide will typically be dynamically selectable, as by use of standard zooming and/or selection tools.

Field 81 of visual display 80 is used to present the output from process 200, that is, to present the bioinformatic prediction of those sequences having the desired function within the genomic sequence. Functional sequences are typically indicated by at least one rectangle 83 (83a, 83b, 83c), the left and right borders of which respectively indicate, by their X-axis coordinates, the starting and ending nucleotides of the region predicted to have function.

Where a single bioinformatic method or approach identifies a plurality of regions having the desired function, a plurality of rectangles 83 is disposed horizontally in field 81. Where multiple methods and/or approaches are used to identify function, each such method and/or approach can be represented by its own series of horizontally disposed rectangles 83, each such horizontally disposed series of rectangles offset vertically from those representing the results of the other methods and approaches.

Thus, rectangles 83a in FIG. 3 represent the functional predictions of a first method of a first approach for predicting function, rectangles 83b represent the functional predictions of a second method and/or second approach for predicting that function, and rectangles 83c represent the predictions of a third method and/or approach.

Where the function desired to be identified is protein coding, field 81 is used to present the bioinformatic prediction of sequences encoding protein.

For example, rectangles 83a can represent the results from GRAIL or GRAIL II, rectangles 83b can represent the results from GENEFINDER, and rectangles 83c can represent the results from DICTION.

Optionally, and preferably, rectangles 83 collectively representing predictions of a single method and/or approach are identically colored and/or textured, and are distinguishable from the color and/or texture used for a different method and/or approach.

Alternatively, or in addition, the color, hue, density, or texture of rectangles 83 can be used further to report a measure of the bioinformatic reliability of the prediction. For example, many gene prediction programs will report a measure of the reliability of prediction. Thus, increasing degrees of such reliability can be indicated, *e.g.*, by increasing density of shading. Where display 80 is used as a graphical user interface, such measures of reliability, and indeed all other results output by the program, can additionally or alternatively be made accessible through linkage from individual rectangles 83, as by time-delayed window ("tool tip" window), or by pointer (*e.g.*, mouse)-activated link.

As earlier described, increased predictive reliability can be achieved by requiring consensus among methods and/or approaches to determining function. Thus, field 81 can include a horizontal series of rectangles 83 that indicate one or more degrees of consensus in predictions of function.

Although FIG. 3 shows three series of horizontally disposed rectangles in field 81, display 80 can include as few as one such series of rectangles and as many as can discriminably be displayed, depending upon the number of methods and/or approaches used to predict a given function.

Furthermore, field 81 can be used to show

predictions of a plurality of different functions. However, the increased visual complexity occasioned by such display makes more useful the ability of the user to select a single function for display. When display 80 is used as a graphical user interface for computer query and analysis, such function can usefully be indicated and user-selectable, as by a series of graphical buttons or tabs (not shown in FIG. 3).

Rectangle 89 is shown in FIG. 3 as including interposed rectangle 84. Rectangle 84 represents the portion of annotated sequence for which predicted functional information has been assayed physically, with the starting and ending nucleotides of the assayed material indicated by the X axis coordinates of the left and right borders of rectangle 84. Rectangle 85, with optional inclusive circles 86 (86a, 86b, and 86c) displays the results of such physical assay.

Although a single rectangle 84 is shown in FIG. 3, physical assay is not limited to just one region of annotated genomic sequence. It is expected that an increasing percentage of regions predicted to have function by process 200 will be assayed physically, and that display 80 will accordingly, for any given genomic sequence, have an increasing number of rectangles 84 and 85, representing an increased density of sequence annotation.

Where the function desired to be identified is protein coding, rectangle 84 identifies the sequence of the probe used to measure expression. In embodiments of the present invention where expression is measured using genome-derived single exon microarrays, rectangle 84 identifies the sequence included within the probe immobilized on the support surface of the microarray. As noted *supra*, such probe will often include a small amount of additional, synthetic, material incorporated during amplification and designed to permit reamplification of the

probe, which sequence is typically not shown in display 80.

Rectangle 87 is used to present the results of bioinformatic assay of the genomic sequence. For example, where the function desired to be identified is protein coding, process 400 can include bioinformatic query of expression databases with the sequences predicted in process 200 to encode exons. And as earlier discussed, because bioinformatic assay presents fewer constraints than does physical assay, often the entire output of process 200 can be used for such assay, without further subsetting thereof by process 300. Therefore, rectangle 87 typically need not have separate indicators therein of regions submitted for bioinformatic assay; that is, rectangle 87 typically need not have regions therein analogous to rectangles 84 within rectangle 89.

Rectangle 87 as shown in FIG. 3 includes smaller rectangles 880 and 88. Rectangles 880 indicate regions that returned a positive result in the bioinformatic assay, with rectangles 88 representing regions that did not return such positive results. Where the function desired to be predicted and displayed is protein coding, rectangles 880 indicate regions of the predicted exons that identify sequence with significant similarity in expression databases, such as EST, SNP, SAGE databases, with rectangles 88 indicating genes novel over those identified in existing expression data bases.

Rectangles 880 can further indicate, through color, shading, texture, or the like, additional information obtained from bioinformatic assay.

For example, where the function assayed and displayed is protein coding, the degree of shading of rectangles 880 can be used to represent the degree of sequence similarity found upon query of expression databases. The number of levels of discrimination can be as few as two (identity, and similarity, where similarity

has a user-selectable lower threshold). Alternatively, as many different levels of discrimination can be indicated as can visually be discriminated.

Where display 80 is used as a graphical user interface, rectangles 880 can additionally provide links directly to the sequences identified by the query of expression databases, and/or statistical summaries thereof. As with each of the precedingly-discussed uses of display 80 as a graphical user interface, it should be understood that the information accessed via display 80 need not be resident on the computer presenting such display, which often will be serving as a client, with the linked information resident on one or more remotely located servers.

Rectangle 85 displays the results of physical assay of the sequence delimited by its left and right borders.

Rectangle 85 can consist of a single rectangle, thus indicating a single assay, or alternatively, and increasingly typically, will consist of a series of rectangles (85a, 85b, 85c) indicating separate physical assays of the same sequence.

Where the function assayed is gene expression, and where gene expression is assayed as herein described using simultaneous two-color fluorescent detection of hybridization to genome-derived single exon microarrays, individual rectangles 85 can be colored to indicate the degree of expression relative to control. Conveniently, shades of green can be used to depict expression in the sample over control values, and shades of red used to depict expression less than control, corresponding to the spectra of the Cy3 and Cy5 dyes conventionally used for respective labeling thereof. Additional functional information can be provided in the form of circles 86 (86a, 86b, 86c), where the diameter of the circle can be used to

indicate expression intensity. As discussed *infra*, such relative expression (expression ratios) and absolute expression (signal intensity) can be expressed using normalized values.

Where display 80 is used as a graphical user interface, rectangle 85 can be used as a link to further information about the assay. For example, where the assay is one for gene expression, each rectangle 85 can be used to link to information about the source of the hybridized mRNA, the identity of the control, raw or processed data from the microarray scan, or the like.

FIG. 4 is rendition of display 80 representing gene prediction and gene expression for a hypothetical BAC, showing conventions used in the Examples presented *infra*. BAC sequence ("Chip seq.") 89 is presented, with the physically assayed region thereof (corresponding to rectangle 84 in FIG. 3) shown in white. Algorithmic gene predictions are shown in field 81, with predictions by GRAIL shown, predictions by GENEFINDER, and predictions by DICTION shown. Within rectangle 87, regions of sequence that, when used to query expression databases, return identical or similar sequences ("EST hit") are shown as white rectangles (corresponding to rectangles 880 in FIG. 3), gray indicates low homology, and black indicates unknowns (where black and gray would correspond to rectangles 88 in FIG. 3).

Although FIGS. 3 and 4 show a single stretch of sequence, uninterrupted from left to right, longer sequences are usefully represented by vertical stacking of such individual Mondrians, as shown in FIGS. 9 and 10.

Single Exon Probes Useful For Measuring Gene Expression

The methods and apparatus of the present invention rapidly produce functional information from

genomic sequence. Where the function to be identified is protein coding, the methods and apparatus of the present invention rapidly identify and confirm the expression of portions of genomic sequence that function to encode protein. As a direct result, the methods and apparatus of the present invention rapidly yield large numbers of single-exon nucleic acid probes, the majority from previously unknown genes, each of which is useful for measuring and/or surveying expression of a specific gene in one or more tissues or cell types.

It is, therefore, another aspect of the present invention to provide genome-derived single exon nucleic acid probes useful for gene expression analysis, and particularly for gene expression analysis by microarray.

Using the methods and genome-derived single-exon microarrays of the present invention, we have for example readily identified a large number of unique ORFs from human genomic sequence. Using single exon probes that encompass these ORFs, we have demonstrated, through microarray hybridization analysis, the expression of 12,673 of these ORFs in Fetal liver.

As would immediately be appreciated by one of skill in the art, each single exon probe having demonstrable expression in Fetal liver is currently available for use in measuring the level of its ORF's expression in Fetal liver.

Diseases of the liver are a significant cause of human morbidity and mortality. Increasingly, genetic factors are being found that contribute to predisposition, onset, and/or aggressiveness of most, if not all, of these diseases; although causative mutations in single genes have been identified for some, these disorders are believed for the most part to have polygenic etiologies.

For example, cirrhosis is a major public health problem. In the industrialized world, it is among the top

ten causes of death; among patients aged 45 to 65, it is the third leading cause of death. The high prevalence is largely the result of alcohol abuse, but other major contributors include chronic hepatitis, biliary disease and iron overload. Approximately 10-15% are cryptogenic.

Cirrhosis is a broad description encompassing the common end stage of many forms of liver injury. Many patients with cirrhosis will remain asymptomatic for years, while others show generalized weakness, anorexia, malaise, and weight loss or, occasionally, more severe symptoms.

The progression from fibrosis, an early consequence of liver disease, to cirrhosis, and the specific histologic morphology that characterizes cirrhosis depend on the extent of injury, the presence of continuing damage, and the response of the liver to damage. The liver may be injured acutely and severely (e.g. necrosis with hepatitis), moderately over months or years (e.g. biliary tract obstruction and chronic active hepatitis), or modestly but continuously (e.g. alcohol abuse).

During the repair process, new vessels connecting the hepatic artery and portal vein to the hepatic venules form within the fibrous sheath that surrounds the surviving nodules of liver cells. These vessels restore the intrahepatic circulatory pathway, but provide relatively low-volume, high-pressure drainage that is less efficient than normal and results in increased portal vein pressure (portal hypertension). Thus, cirrhosis is not static and its features depend on the disease activity and stage.

As cirrhosis is the end stage of many forms of liver disease, many genes have been identified that can contribute to the development of cirrhosis. These include, e.g., the genes responsible for Wilson disease (Online Mendelian Inheritance of Man ("OMIM") 277900), type IV glycogen storage disease (OMIM 232500), galactosemia (OMIM

230400), and a deficiency of alpha-1-antitrypsin (OMIM 107400). There is substantial evidence, however, for as yet uncharacterized loci which cause cirrhosis.

For example, Iber and Maddrey, Prog. Liver Dis. 2: 290-302 (1965), reviewed 13 previously reported families and 8 new to this study, each with 2 or more affected members. They pointed out that, with a single exception, the multiple cases were in the same generation. Within a given family, the age of onset, clinical course, and biopsy findings were very similar, but there were wide differences between families.

Kalra et al., Hum. Hered. 32:170-175 (1982) studied the families of 220 cases of Indian childhood cirrhosis and 70 families of age-matched controls. The hypotheses of autosomal recessive, partial sex-linkage, and doubly recessive inheritance were found untenable and the authors concluded that multifactorial inheritance was most plausible. Lefkowitch et al., New Eng. J. Med. 307:271-277 (1982) described 4 white American sibs who died between ages 4.5 and 6 years of cirrhosis that closely resembled that of the childhood cirrhosis of Asiatic Indians.

Another example of uncharacterized loci which cause cirrhosis are those related to the risk of alcoholism.

Cloninger, Science 236:410-416 (1987), defined two separate types of alcoholism. According to these definitions, type 1 alcohol abuse has its usual onset after the age of 25 years and is characterized by severe psychological dependence and guilt. Type 1 occurs in both men and women and requires both genetic and environmental factors to become manifest. By contrast, type 2 alcohol abuse has its onset before the age of 25; persons with this type of alcoholism are characterized by their inability to abstain from alcohol and by frequent aggressive and antisocial behavior. Type 2 alcoholism is rarely found in

women and is much more heritable.

Despite considerable effort to identify genes related to the risk of alcoholism, relatively few genes have been identified. Some of this work has suggested a relationship between the metabolism of dopamine and alcoholism. Blum et al., J.A.M.A. 263:2055-2060 (1990) and Bolos et al., J.A.M.A. 264:3156-3160 (1990) investigated the relationship of the dopamine D2 receptor (DRD2; OMIM 126450) to alcoholism, but the sample size was small and their results were inconclusive. However, Tiihonen et al., Molec. Psychiat. 4, 286-289 (1999), found a markedly higher frequency in a population of type 1 alcoholics of the low activity allele of the enzyme catechol-O-methyltransferase (COMT, OMIM 116790), which has a crucial role in the metabolism of dopamine, suggesting a role for dopamine metabolism in increased risk of alcoholism. For a brief review of recent progress toward the identification of genes related to risk for alcoholism see Buck, Genome 9:927-928 (1998).

As another example, multiple genes have been shown to predispose to hyperlipoproteinemia or hyperlipidemia. Much attention has been focused on these disorders because there is a strong association of hyperlipidemia, especially hypercholesterolemia, with development of coronary artery disease. Coronary artery disease accounts for at least 25% of all deaths in the United States. Coronary artery disease results when the arteries supplying the heart muscle become occluded by plaques composed of lipids like cholesterol, blood clotting components and blood cells.

The major plasma lipids circulate bound to proteins as macromolecular complexes called lipoproteins. Although closely interrelated, the major lipoprotein classes - chylomicron, very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL), and high-density lipoprotein

(HDL) - are usually classified in terms of physicochemical properties (e.g., density after centrifugation). Chylomicrons, the largest lipoproteins, carry exogenous triglyceride from the intestine via the thoracic duct to the venous system and into peripheral sites. VLDL carries endogenous triglyceride primarily from the liver to the same peripheral sites for storage or use. Lipases quickly degrade the triglyceride in VLDL to produce intermediate density lipoproteins (IDL) and within 2 to 6 h, IDL is degraded further to generate LDL, which has a plasma half-life of 2 to 3 days. While the overall fate of LDL is unclear, the liver is responsible for removing approximately 70% and active receptor sites have been found on the surfaces of hepatocytes.

Several monogenic conditions that lead to elevated levels of one or more serum lipoproteins have been defined and the responsible gene identified, including, e.g., hyperlipoproteinemia type I (OMIM 238600), familial hypercholesterolemia (OMIM 143890), and familial defective apolipoprotein B (OMIM 107730). However, in many cases the etiology is unknown and there is strong evidence for additional uncharacterized loci.

For example, Zuliani et al., Arterioscler. Thromb. Vasc. Biol. 19:802-809 (1999) identified a Sardinian family with a recessive form of hypercholesterolemia with the clinical features of familial hypercholesterolemia (OMIM 603813), and found that previously identified genes were not responsible for this disorder. They proposed that in this new lipid disorder, a recessive defect causes a selective impairment of the LDL receptor function in the liver. Ciccarese et al., Am. J. Hum. Genet. 66:453-460 (2000) recently mapped this novel disease locus.

Another example is designated familial combined hyperlipidemia (OMIM 144250) which affects approximately 1-

2% of the population in the Western world. This disorder can have its basis in mutation in several novel genes, two of which have been mapped to chromosome 1 (Pajukanta et al., Nature Genet. 18:369-373 (1998)) and chromosome 11 (Aouizerat et al., Am. J. Hum. Genet. 65, 397-412 (1999)). The high frequency of this disorder suggests that most, if not all, hyperlipidemias are of multifactorial genetic etiology.

As yet a further example, primary schlerosing cholangitis (PSC) is a disorder characterized by a patchy obliterative inflammatory fibrosis of the large bile ducts. Chronic inflammation leads to extensive bile duct strictures, cholestasis, and gradual progression to biliary cirrhosis. PSC occurs most often in young men and is commonly associated with inflammatory bowel disease, especially ulcerative colitis. The onset is usually insidious, with gradual, progressive fatigue, pruritus, and jaundice. There is no specific therapy for sclerosing cholangitis, and liver transplantation is the only apparent cure.

The etiology of PSC is not known, but both genetic and immunologic abnormalities have been implicated. However, the frequency of HLA-B8 and HLA-DT2, which are associated with a number of autoimmune diseases, is higher in PSC than normal individuals. Prochazka et al., New Eng. J. Med. 322:1842-1844 (1990) found that 100% of 29 patients with primary sclerosing cholangitis carried the HLA-DRw52a antigen, which is normally present in 35% of the population.

As a still further example, sarcoidosis is a disease of unknown cause characterized by non-caseating granulomas in one or more organ systems. These granulomas may resolve completely or proceed to fibrosis. The disorder is systemic, but the liver is affected in approximately 75% of cases. Sarcoidosis occurs mainly in persons aged 20 to

40 yr and is most common in Northern Europeans and American blacks. The lifetime risk of developing sarcoidosis is particularly high among Swedish men (1.15%), Swedish women (1.6%), and African Americans (2.4%).

The much greater frequency in African Americans relative to the United States population overall suggests a genetic contribution to etiology. Early research studying familial aggregation indicated that the disease may have a nongenetic basis because the family pattern did not conform to a simple Mendelian mode of inheritance (Allison, Sth. Med. J. 57: 27-32 (1964)). However, Headings et al., Ann. N.Y. Acad. Sci. 278:377-385 (1976) favored multifactorial genetic inheritance of susceptibility. Nowack et al., Arch. Intern. Med. 147:481-483 (1987), found an unusually high frequency of HLA-DR5 in a study of 440 patients with sarcoidosis in Marburg, Germany. They also concluded that the role of an environmental or infectious agent triggering sarcoidosis cannot be envisaged without considering genetically linked cofactors.

Other significant diseases of liver are also believed to have a genetic, typically polygenic, etiologic component. These diseases include, e.g., primary biliary cirrhosis, Zellweger syndrome, cholestasis-lymphedema syndrome, Alstrom syndrome, primary pulmonary hypertension, Berardinelli-Seip congenital lipodystrophy, iron overload in Africa, neonatal cholestatic hepatitis, autosomal recessive KID syndrome, familial hypotransferrinemia, type I congenital dyserythropoietic anemia, porphyria variegata, Finnish lactic acidosis with hepatic hemosiderosis, Rotor syndrome, essential hypertension, ARC syndrome, type II conjugated hyperbilirubinemia, Lambert syndrome, ichthyosis congenita with biliary atresia, Kabuki make-up syndrome, Meckel syndrome, cerebral aneurysm-cirrhosis syndrome, glycogen storage diseases, polycystic kidney and hepatic disease,

isolated Caroli disease, trisomy 18-like syndrome, Osler-Rendu-Weber syndrome 3, fatal intrahepatic cholestasis, Coach syndrome, type C Niemann-Pick disease, hereditary persistence of fetal hemoglobin and hepatocellular cancer.

Altered responses to a variety of infectious agents that target the liver, especially acute viral hepatitis, have also been shown or are suspected to have genetic bases or contributions. In addition to differential susceptibility to primary infectious agents, these altered responses include predisposition to complicating conditions following contact with particular infectious agents. These include, e.g., development of hepatocellular carcinoma 2 correlated with Hepatitis B infection, and severe hepatic fibrosis following Schistosoma mansoni infection.

The central role of the liver in drug metabolism results in exposure of this organ to a large variety of potentially toxic chemical agents and metabolites. These include naturally occurring plant alkaloids and mycotoxins, industrial chemicals, and, additionally, pharmacologic agents used in treating disease. The range of manifestations of toxin- and drug-induced liver disease are virtually as broad as the range of acute and chronic disorders and have also been shown or suspected to have genetic bases or contributions.

Such interactions between drugs and genotype have been shown in the response, e.g., to the anticonvulsant phenytoin, which can cause severe hepatitis-like disease in individuals who are impaired in the ability to detoxify a metabolite of phenytoin in the liver, and in the response to the drug sodium valproate, which can produce severe hepatotoxicity in certain individuals. The abnormal responses to both of these drugs are believed to be influenced by underlying genetic factors.

The human genome-derived single exon nucleic acid

probes and microarrays of the present invention are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of human liver, particularly those diseases with polygenic etiology. With each of the single exon probes described herein shown to be expressed at detectable levels in human liver, and with about 2/3 of the probes identifying novel genes, the single exon microarrays of the present invention provide exceptionally high informational content for such studies.

For example, diagnosis (including differential diagnosis among clinically indistinguishable disorders, such as cirrhosis), staging, and/or grading of a disease can be based upon the quantitative relatedness of a patient gene expression profile to one or more reference expression profiles known to be characteristic of a given liver disease, or to specific grades or stages thereof.

In one embodiment, the patient gene expression profile is generated by hybridizing nucleic acids obtained directly or indirectly from transcripts expressed in the patient's liver to the genome-derived single exon microarray of the present invention. Reference profiles are obtained similarly, using nucleic acids obtained directly or indirectly from transcripts expressed by liver of individuals with known liver disease. Methods for quantitatively relating gene expression profiles, without regard to the function of the protein encoded by the gene, are disclosed in WO 99/58720, incorporated herein by reference in its entirety.

In another approach, the genome-derived single exon probes and microarrays of the present invention can be used to interrogate genomic DNA, rather than pools of expressed message; this latter approach permits predisposition to and/or prognosis of liver disease to be assessed through the massively parallel determination of altered copy number, deletion, or mutation in the patient's

genome of exons known to be expressed in human liver. The algorithms set forth in WO 99/58720 can be applied to such genomic profiles without regard to the function of the protein encoded by the interrogated gene.

The utility is specific to the probe; at sufficiently high hybridization stringency, which stringencies are well known in the art – see Ausubel et al. and Maniatis *et al.* – each probe reports the level of expression of message specifically containing that ORF.

It should be appreciated, however, that the probes of the present invention, for which expression in the Fetal liver has been demonstrated are useful for both measurement in the Fetal liver and for survey of expression in other tissues.

Significant among such advantages is the presence of probes for novel genes.

As mentioned above and further detailed in Examples 1 and 2, the methods described enable ORFs which are not present in existing expression databases to be identified. And the fewer the number of tissues in which the ORF can be shown to be expressed, the more likely the ORF will prove to be part of a novel gene: as further discussed in Example 2, ORFs whose expression was measurable in only a single of the tested tissues were represented in existing expression databases at a rate of only 11%, whereas 36% of ORFs whose expression was measurable in 9 tissues were present in existing expression databases, and fully 45% of those ORFs expressed in all ten tested tissues were present in existing expressed sequence databases.

Either as tools for measuring gene expression or tools for surveying gene expression, the genome-derived single exon probes of the present invention have significant advantages over the cDNA or EST-based probes that are currently available for achieving these utilities.

The genome-derived single exon probes of the present invention are useful in constructing genome-derived single exon microarrays; the genome-derived single exon microarrays, in turn, are useful devices for measuring and for surveying gene expression in the human.

Gene expression analysis using microarrays – conventionally using microarrays having probes derived from expressed message – is well-established as useful in the biological research arts (see Lockhart et al. Nature 405, 827-836).

Microarrays have been used to determine gene expression profiles in cells in response to drug treatment (see, for example, Kaminski *et al.*, "Global Analysis of Gene Expression in Pulmonary Fibrosis Reveals Distinct Programs Regulating Lung Inflammation and Fibrosis," *Proc. Natl. Acad. Sci. USA* 97(4):1778-83 (2000); Bartosiewicz *et al.*, "Development of a Toxicological Gene Array and Quantitative Assessment of This Technology," *Arch. Biochem. Biophys.* 376(1):66-73 (2000)), viral infection (see for example, Geiss *et al.*, "Large-scale Monitoring of Host Cell Gene Expression During HIV-1 Infection Using cDNA Microarrays," *Virology* 266(1):8-16 (2000)) and during cell processes such as differentiation, senescence and apoptosis (see, for example, Shelton *et al.*,"Microarray Analysis of Replicative Senescence," *Curr. Biol.* 9(17):939-45 (1999); Voehringer *et al.*, "Gene Microarray Identification of Redox and Mitochondrial Elements That Control Resistance or Sensitivity to Apoptosis," Proc. Natl. Acad. Sci. USA 97(6):2680-5 (2000)).

Microarrays have also been used to determine abnormal gene expression in diseased tissues (see, for example, Alon *et al.*, "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays," *Proc. Natl. Acad. Sci. USA* 96(12):6745-50 (1999); Perou *et al.*,

"Distinctive Gene Expression Patterns in Human Mammary Epithelial Cells and Breast Cancers, *Proc. Natl. Acad. Sci. USA* 96(16):9212-7 (1999); Wang *et al.*, "Identification of Genes Differentially Over-expressed in Lung Squamous Cell Carcinoma Using Combination of cDNA Subtraction and Microarray Analysis," *Oncogene* 19(12):1519-28 (2000); Whitney *et al.*, "Analysis of Gene Expression in Multiple Sclerosis Lesions Using cDNA Microarrays," *Ann. Neurol.* 46(3):425-8 (1999)), in drug discovery screens (see, for example, Scherf *et al.*, "A Gene Expression Database for the Molecular Pharmacology of Cancer," *Nat. Genet.* 24(3):236-44 (2000)) and in diagnosis to determine appropriate treatment strategies (see, for example, Sgroi *et al.*, "*In vivo* Gene Expression Profile Analysis of Human Breast Cancer Progression," *Cancer Res.* 59(22):5656-61 (1999)).

In microarray-based gene expression screens of pharmacological drug candidates upon cells, each probe provides specific useful data. In particular, it should be appreciated that even those probes that show no change in expression are as informative as those that do change, serving, in essence, as negative controls.

For example, where gene expression analysis is used to assess toxicity of chemical agents on cells, the failure of the agent to change a gene's expression level is evidence that the drug likely does not affect the pathway of which the gene's expressed protein is a part. Analogously, where gene expression analysis is used to assess side effects of pharmacological agents – whether in lead compound discovery or in subsequent screening of lead compound derivatives – the inability of the agent to alter a gene's expression level is evidence that the drug does not affect the pathway of which the gene's expressed protein is a part.

WO 99/58720 provides methods for quantifying the relatedness of a first and second gene expression profile

and for ordering the relatedness of a plurality of gene expression profiles. The methods so described permit useful information to be extracted from a greater percentage of the individual gene expression measurements from a microarray than methods previously used in the art.

Other uses of microarrays are described in Gerhold *et al.*, *Trends Biochem. Sci.* 24(5):168-173 (1999) and Zweiger, *Trends Biotechnol.* 17(11):429-436 (1999); Schena et al.

The invention particularly provides genome-derived single-exon probes known to be expressed in Fetal liver. The individual single exon probes can be provided in the form of substantially isolated and purified nucleic acid, typically, but not necessarily, in a quantity sufficient to perform a hybridization reaction.

Such nucleic acid can be in any form directly hybridizable to the message that contains the probe's ORF, such as double stranded DNA, single-stranded DNA complementary to the message, single-stranded RNA complementary to the message, or chimeric DNA/RNA molecules so hybridizable. The nucleic acid can alternatively or additionally include either nonnative nucleotides, alternative internucleotide linkages, or both, so long as complementary binding can be obtained. For example, probes can include phosphorothioates, methylphosphonates, morpholino analogs, and peptide nucleic acids (PNA), as are described, for example, in U.S. Patent Nos. 5,142,047; 5,235,033; 5,166,315; 5,217,866; 5,184,444; 5,861,250.

Usefully, however, such probes are provided in a form and quantity suitable for amplification, where the amplified product is thereafter to be used in the hybridization reactions that probe gene expression. Typically, such probes are provided in a form and quantity suitable for amplification by PCR or by other well known amplification technique. One such technique additional to

PCR is rolling circle amplification, as is described, *inter alia*, in U.S. Patent Nos. 5,854,033 and 5,714,320 and international patent publications WO 97/19193 and WO 00/15779. As is well understood, where the probes are to be provided in a form suitable for amplification, the range of nucleic acid analogues and/or internucleotide linkages will be constrained by the requirements and nature of the amplification enzyme.

Where the probe is to be provided in form suitable for amplification, the quantity need not be sufficient for direct hybridization for gene expression analysis, and need be sufficient only to function as an amplification template, typically at least about 1, 10 or 100 pg or more.

Each discrete amplifiable probe can also be packaged with amplification primers, either in a single composition that comprises probe template and primers, or in a kit that comprises such primers separately packaged therefrom. As earlier mentioned, the ORF-specific 5' primers used for genomic amplification can have a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification can have a second, different, common sequence added thereto, thus permitting, in this embodiment, the use of a single set of 5' and 3' primers to amplify any one of the probes. The probe composition and/or kit can also include buffers, enzyme, *etc.*, required to effect amplification.

As mentioned earlier, when intended for use on a genome-derived single exon microarray of the present invention, the genome-derived single exon probes of the present invention will typically average at least about 100, 200, 300, 400 or 500 bp in length, including (and typically, but not necessarily centered about) the ORF. Furthermore, when intended for use on a genome-derived single exon microarray of the present invention, the

genome-derived single exon probes of the present invention will typically not contain a detectable label.

When intended for use in solution phase hybridization, however – that is, for use in a hybridization reaction in which the probe is not first bound to a support substrate (although the target may indeed be so bound) – length constraints that are imposed in microarray-based hybridization approaches will be relaxed, and such probes will typically be labeled.

In such case, the only functional constraint that dictates the minimum size of such probe is that each such probe must be capable of specifically identifying in a hybridization reaction the exon from which it is drawn. In theory, a probe of as little as 17 nucleotides is capable of uniquely identifying its cognate sequence in the human genome. For hybridization to expressed message – a subset of target sequence that is much reduced in complexity as compared to genomic sequence – even fewer nucleotides are required for specificity.

Therefore, the probes of the present invention can include as few as 20, 25 or 50 bp or ORF, or more. In particular embodiments, the ORF sequences are given in SEQ ID NOS. 12,674 - 25,129, respectively, for probe SEQ ID NOS. 1 - 12,673. The minimum amount of ORF required to be included in the probe of the present invention in order to provide specific signal in either solution phase or microarray-based hybridizations can readily be determined for each of ORF SEQ ID NOS. 12,674 - 25,129 individually by routine experimentation using standard high stringency conditions.

Such high stringency conditions are described, *inter alia*, in Ausubel et al. and Maniatis *et al*. For microarray-based hybridization, standard high stringency conditions can usefully be 50% formamide, 5X SSC, 0.2 µg/µl poly(dA), 0.2 µg/µl human c_ot1 DNA, and 0.5 % SDS, in a

humid oven at 42°C overnight, followed by successive washes of the microarray in 1X SSC, 0.2% SDS at 55°C for 5 minutes, and then 0.1X SSC, 0.2% SDS, at 55°C for 20 minutes. For solution phase hybridization, standard high stringency conditions can usefully be aqueous hybridization at 65°C in 6X SSC. Lower stringency conditions, suitable for cross-hybridization to mRNA encoding structurally- and functionally-related proteins, can usefully be the same as the high stringency conditions but with reduction in temperature for hybridization and washing to room temperature (approximately 25°C).

When intended for use in solution phase hybridization, the maximum size of the single exon probes of the present invention is dictated by the proximity of other expressed exons in genomic DNA: although each single exon probe can include intergenic and/or intronic material contiguous to the ORF in the human genome, each probe of the present invention will include portions of only one expressed exon.

Thus, each single exon probe will include no more than about 25 kb of contiguous genomic sequence, more typically no more than about 20 kb of contiguous genomic sequence, more usually no more than about 15 kb, even more usually no more than about 10 kb. Usually, probes that are maximally about 5 kb will be used, more typically no more than about 3 kb.

It will be appreciated that the Sequence Listing appended hereto presents, by convention, only that strand of the probe and ORF sequence that can be directly translated reading from 5' to 3' end. As would be well understood by one of skill in the art, single stranded probes must be complementary in sequence to the ORF as present in an mRNA; it is well within the skill in the art to determine such complementary sequence. It will further be understood that double stranded probes can be used in

both solution-phase hybridization and microarray-based hybridization if suitably denatured.

Thus, it is an aspect of the present invention to provide single-stranded nucleic acid probes that have sequence complementary to those described herein above and below, and double-stranded probes one strand of which has sequence complementary to the probes described herein.

The probes can, but need not, contain intergenic and/or intronic material that flanks the ORF, on one or both sides, in the same linear relationship to the ORF that the intergenic and/or intronic material bears to the ORF in genomic DNA. The probes do not, however, contain nucleic acid derived from more than one expressed ORF.

And when intended for use in solution hybridization, the probes of the present invention can usefully have detectable labels. Nucleic acid labels are well known in the art, and include, *inter alia*, radioactive labels, such as ^{3}H, ^{32}P, ^{33}P, ^{35}S, ^{125}I, ^{131}I; fluorescent labels, such as Cy3, Cy5, Cy5.5, Cy7, SYBR®

Green and other labels described in Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 7th ed., Molecular Probes Inc., Eugene, OR (2000), or fluorescence resonance energy transfer tandem conjugates thereof; labels suitable for chemiluminescent and/or enhanced chemiluminescent detection; labels suitable for ESR and NMR detection; and labels that include one member of a specific binding pair, such as biotin, digoxigenin, or the like.

The probes, either in quantity sufficient for hybridization or sufficient for amplification, can be provided in individual vials or containers.

Alternatively, such probes can usefully be packaged as a plurality of such individual genome-derived single exon probes.

When provided as a collection of plural

individual probes, the probes are typically made available in amplifiable form in a spatially-addressable ordered set, typically one per well of a microtiter dish. Although a 96 well microtiter plate can be used, greater efficiency is obtained using higher density arrays.

If, as earlier mentioned, the ORF-specific 5' primers used for genomic amplification had a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification had a second, different, common sequence added thereto, a single set of 5' and 3' primers can be used to amplify all of the probes from the amplifiable ordered set.

Such collections of genome-derived single exon probes can usefully include a plurality of probes chosen for the common attribute of expression in the human Fetal liver.

In such defined subsets, typically at least 50, 60, 75, 80, 85, 90 or 95% or more of the probes will be chosen by their expression in the defined tissue or cell type.

The single exon probes of the present invention, as well as fragments of the single exon probes comprising selectively hybridizable portions of the probe ORF, can be used to obtain the full length cDNA that includes the ORF by (i) screening of cDNA libraries; (ii) rapid amplification of cDNA ends ("RACE"); or (iii) other conventional means, as are described, *inter alia*, in Ausubel et al. and Maniatis *et al*.

It is another aspect of the present invention to provide genome-derived single exon nucleic acid microarrays useful for gene expression analysis, where the term "microarray" has the meaning given in the definitional section of this description, *supra*.

The invention particularly provides genome-derived single-exon nucleic acid microarrays comprising a

plurality of probes known to be expressed in human Fetal liver. In preferred embodiments, the present invention provides human genome-derived single exon microarrays comprising a plurality of probes drawn from the group consisting of SEQ ID NOS.: 1 - 12,673.

When used for gene expression analysis, the genome-derived single exon microarrays provide greater physical informational density than do the genome-derived single exon microarrays that have lower percentages of probes known to be expressed commonly in the tested tissue. At a fixed probe density, for example, a given microarray surface area of the defined subset genome-derived single exon microarray can yield a greater number of expression measurements. Alternatively, at a given probe density, the same number of expression measurements can be obtained from a smaller substrate surface area. Alternatively, at a fixed probe density and fixed surface area, probes can be provided redundantly, providing greater reliability in signal measurement for any given probe. Furthermore, with a higher percentage of probes known to be expressed in the assayed tissue, the dynamic range of the detection means can be adjusted to reveal finer levels discrimination among the levels of expression.

Although particularly described with respect to their utility as probes of gene expression, particularly as probes to be included on a genome-derived single exon microarray, each of the nucleic acids having SEQ ID NOS.: 1 - 12,673 contains an open-reading frame, set forth respectively in SEQ ID NOS.: 12,674 - 25,129, that encodes a protein domain. Thus, each of SEQ ID NOS. 1 - 12,673 can be used, or that portion thereof in SEQ ID NOS. 12,674 - 25,129 used, to express a protein domain by standard *in vitro* recombinant techniques. See Ausubel et al. and Maniatis *et al.*

Additionally, kits are available commercially

that readily permit such nucleic acids to be expressed as protein in bacterial cells, insect cells, or mammalian cells, as desired (*e.g.*, HAT™ Protein Expression & Purification System, ClonTech Laboratories, Palo Alto, CA; Adeno-X™ Expression System, ClonTech Laboratories, Palo Alto, CA; Protein Fusion & Purification (pMAL™) System, New England Biolabs, Beverley, MA)

Furthermore, shorter peptides can be chemically synthesized using commercial peptide synthesizing equipment and well known techniques. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series, (Paper)), Oxford Univ. Press (March 2000) (ISBN: 0199637245); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7) , Oxford Univ. Press (August 1992) (ISBN: 0198556683); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (December 1993) (ISBN: 0387564314).

It is, therefore, another aspect of the invention to provide peptides comprising an amino acid sequence translated from SEQ ID NOS.: 12,674 - 25,129. Such amino acid sequences are set out in SEQ ID NOS: 25,130 - 37,156. Any such recombinantly-expressed or synthesized peptide of at least 8, and preferably at least about 15, amino acids, can be conjugated to a carrier protein and used to generate antibody that recognizes the peptide. Thus, it is a further aspect of the invention to provide peptides that have at least 8, preferably at least 15, consecutive amino acids.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

Preparation of Single Exon Microarrays from ORFs Predicted

in Human Genomic Sequence

Bioinformatics Results

All human BAC sequences in fewer than 10 pieces that had been accessioned in a five month period immediately preceding this study were downloaded from GenBank. This corresponds to ~2200 clones, totaling ~350 MB of sequence, or approximately 10% of the human genome.

After masking repetitive elements using the program CROSS_MATCH, the sequence was analyzed for open reading frames using three separate gene finding programs. The three programs predict genes using independent algorithmic methods developed on independent training sets: GRAIL uses a neural network, GENEFINDER uses a hidden Markoff model, and DICTION, a program proprietary to Genetics Institute, operates according to a different heuristic. The results of all three programs were used to create a prediction matrix across the segment of genomic DNA.

The three gene finding programs yielded a range of results. GRAIL identified the greatest percentage of genomic sequence as putative coding region, 2% of the data analyzed. GENEFINDER was second, calling 1%, and DICTION yielded the least putative coding region, with 0.8% of genomic sequence called as coding region.

The consensus data were as follows. GRAIL and GENEFINDER agreed on 0.7% of genomic sequence, GRAIL and DICTION agreed on 0.5% of genomic sequence, and the three programs together agreed on 0.25% of the data analyzed. That is, 0.25% of the genomic sequence was identified by all three of the programs as containing putative coding region.

ORFs predicted by any two of the three programs ("consensus ORFs") were assorted into "gene bins" using two criteria: (1) any 7 consecutive exons within a 25 kb window

were placed together in a bin as likely contributing to a single gene, and (2) all ORFs within a 25 kb window were placed together in a bin as likely contributing to a single gene if fewer than 7 exons were found within the 25 kb window.

PCR

The largest ORF from each gene bin that did not span repetitive sequence was then chosen for amplification, as were all consensus ORFs longer than 500 bp. This method approximated one exon per gene; however, a number of genes were found to be represented by multiple elements.

Previously, we had determined that DNA fragments fewer than 250 bp in length do not bind well to the amino-modified glass surface of the slides used as support substrate for construction of microarrays; therefore, amplicons were designed in the present experiments to approximate 500 bp in length.

Accordingly, after selecting the largest ORF per gene bin, a 500 bp fragment of sequence centered on the ORF was passed to the primer picking software, PRIMER3 (available online for use at http://www-genome.wi.mit.edu/cgi-bin/primer/). A first additional sequence was commonly added to each ORF-unique 5' primer, and a second, different, additional sequence was commonly added to each ORF-unique 3' primer, to permit subsequent reamplification of the amplicon using a single set of "universal" 5' and 3' primers, thus immortalizing the amplicon. The addition of universal priming sequences also facilitates sequence verification, and can be used to add a cloning site should some ORFs be found to warrant further study.

The ORFs were then PCR amplified from genomic DNA, verified on agarose gels, and sequenced using the universal primers to validate the identity of the amplicon

to be spotted in the microarray.

Primers were supplied by Operon Technologies (Alameda, CA). PCR amplification was performed by standard techniques using human genomic DNA (Clontech, Palo Alto, CA) as template. Each PCR product was verified by SYBR® green (Molecular Probes, Inc., Eugene, OR) staining of agarose gels, with subsequent imaging by Fluorimager (Molecular Dynamics, Inc., Sunnyvale, CA). PCR amplification was classified as successful if a single band appeared.

The success rate for amplifying ORFs of interest directly from genomic DNA using PCR was approximately 75%. FIG. 5 graphs the distribution of predicted ORF (exon) length and distribution of amplified PCR products, with ORF length shown in red and PCR product length shown in blue (which may appear black in the figure). Although the range of ORF sizes is readily seen to extend to beyond 900 bp, the mean predicted exon size was only 229 bp, with a median size of 150 bp (n=9498). With an average amplicon size of 475 ± 25 bp, approximately 50% of the average PCR amplification product contained predicted coding region, with the remaining 50% of the amplicon containing either intron, intergenic sequence, or both.

Using a strategy predicated on amplifying about 500 bp, it was found that long exons had a higher PCR failure rate. To address this, the bioinformatics process was adjusted to amplify 1000, 1500 or 2000 bp fragments from exons larger than 500 bp. This improved the rate of successful amplification of exons exceeding 500 bp, constituting about 9.2% of the exons predicted by the gene finding algorithms.

Approximately 75% of the probes disposed on the array (90% of those that successfully PCR amplified) were sequence-verified by sequencing in both the forward and reverse direction using MegaBACE sequencer (Molecular

Dynamics, Inc., Sunnyvale, CA), universal primers, and standard protocols.

Some genomic clones (BACs) yielded very poor PCR and sequencing results. The reasons for this are unclear, but may be related to the quality of early draft sequence or the inclusion of vector and host contamination in some submitted sequence data.

Although the intronic and intergenic material flanking coding regions could theoretically interfere with hybridization during microarray experiments, subsequent empirical results demonstrated that differential expression ratios were not significantly affected by the presence of noncoding sequence. The variation in exon size was similarly found not to affect differential expression ratios significantly; however, variation in exon size was observed to affect the absolute signal intensity (data not shown).

The 350 MB of genomic DNA was, by the above-described process, reduced to 9750 discrete probes, which were spotted in duplicate onto glass slides using commercially available instrumentation (MicroArray GenII Spotter and/or MicroArray GenIII Spotter, Molecular Dynamics, Inc., Sunnyvale, CA). Each slide additionally included either 16 or 32 *E. coli* genes, the average hybridization signal of which was used as a measure of background biological noise.

Each of the probe sequences was BLASTed against the human EST data set, the NR data set, and SwissProt GenBank (May 7, 1999 release 2.0.9).

One third of the probe sequences (as amplified) produced an exact match (BLAST Expect ("E") values less than 1 e^{-100}) to either an EST (20% of sequences) or a known mRNA (13% of sequences). A further 22% of the probe sequences showed some homology to a known EST or mRNA (BLAST E values from 1 e^{-5} to 1 e^{-99}). The remaining 45% of

the probe sequences showed no significant sequence homology to any expressed, or potentially expressed, sequences present in public databases.

All of the probe sequences (as amplified) were then analyzed for protein similarities with the SwissProt database using BLASTX, Gish *et al.*, *Nature Genet.* 3:266 (1993). The predicted functional breakdowns of the 2/3 of probes identical or homologous to known sequences are presented in Table 1.

Table 1

Function of Predicted ORFs As Deduced From Comparative Sequence Analysis			
Total	V6 chip	V7 chip	Function Predicted from Comparative Sequence Analysis
211	96	115	Receptor
120	43	77	Zinc Finger
30	11	19	Homeobox
25	9	16	Transcription Factor
17	11	7	Transcription
118	57	61	Structural
95	39	56	Kinase
36	18	18	Phosphatase
83	31	52	Ribosomal
45	19	26	Transport
21	17	14	Growth Factor
17	12	5	Cytochrome
50	33	17	Channel

As can be seen, the two most common types of genes were transcription factors and receptors, making up 2.2% and 1.8% of the arrayed elements, respectively.

EXAMPLE 2

Gene Expression Measurements From Genome-Derived Single Exon Microarrays

The two genome-derived single exon microarrays prepared according to Example 1 were hybridized in a series of simultaneous two-color fluorescence experiments to (1) Cy3-labeled cDNA synthesized from message drawn individually from each of brain, heart, liver, fetal liver, placenta, lung, bone marrow, HeLa, BT 474, or HBL 100 cells, and (2) Cy5-labeled cDNA prepared from message pooled from all ten tissues and cell types, as a control in each of the measurements. Hybridization and scanning were carried out using standard protocols and Molecular Dynamics equipment.

Briefly, mRNA samples were bought from commercial sources (Clontech, Palo Alto, CA and Amersham Pharmacia Biotech (APB)). Cy3-dCTP and Cy5-dCTP (both from APB) were incorporated during separate reverse transcriptions of 1 µg of polyA$^+$ mRNA performed using 1 µg oligo(dT)12-18 primer and 2 µg random 9mer primers as follows. After heating to 70°C, the RNA:primer mixture was snap cooled on ice. After snap cooling on ice, added to the RNA to the stated final concentration was: 1X Superscript II buffer, 0.01 M DTT, 100µM dATP, 100 µM dGTP, 100 µM dTTP, 50 µM dCTP, 50 µM Cy3-dCTP or Cy5-dCTP 50 µM, and 200 U Superscript II enzyme. The reaction was incubated for 2 hours at 42°C. After 2 hours, the first strand cDNA was isolated by adding 1 U Ribonuclease H, and incubating for 30 minutes at 37°C. The reaction was then purified using a Qiagen PCR cleanup column, increasing the number of ethanol washes to 5. Probe was eluted using 10 mM Tris pH 8.5.

Using a spectrophotometer, probes were measured for dye incorporation. Volumes of both Cy3 and Cy5 cDNA corresponding to 50 pmoles of each dye were then dried in a Speedvac, resuspended in 30 µl hybridization solution containing 50% formamide, 5X SSC, 0.2 µg/µl poly(dA), 0.2 µg/µl human c_0t1 DNA, and 0.5 % SDS.

Hybridizations were carried out under a coverslip, with the array placed in a humid oven at 42°C overnight. Before scanning, slides were washed in 1X SSC, 0.2% SDS at 55°C for 5 minutes, followed by 0.1X SSC, 0.2% SDS, at 55°C for 20 minutes. Slides were briefly dipped in water and dried thoroughly under a gentle stream of nitrogen.

Slides were scanned using a Molecular Dynamics Gen3 scanner, as described. Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376).

Although the use of pooled cDNA as a reference permitted the survey of a large number of tissues, it attenuates the measurement of relative gene expression, since every highly expressed gene in the tissue/cell type-specific fluorescence channel will be present to a level of at least 10% in the control channel. Because of this fact, both signal and expression ratios (the latter hereinafter, "expression" or "relative expression") for each probe were normalized using the average ratio or average signal, respectively, as measured across the whole slide.

Data were accepted for further analysis only when signal was at least three times greater than biological noise, the latter defined by the average signal produced by the *E. coli* control genes.

The relative expression signal for these probes was then plotted as function of tissue or cell type, and is presented in FIG. 6.

FIG. 6 shows the distribution of expression across a panel of ten tissues. The graph shows the number of sequence-verified products that were either not expressed ("0"), expressed in one or more but not all tested tissues ("1" - "9"), and expressed in all tissues tested ("10").

Of 9999 arrayed elements on the two microarrays (including positive and negative controls and "failed" products), 2353 (51%) were expressed in at least one tissue or cell type. Of the gene elements showing significant signal – where expression was scored as "significant" if the normalized Cy3 signal was greater than 1, representing signal 5-fold over biological noise (0.2) – 39% (991) were expressed in all 10 tissues. The next most common class (15%) consisted of gene elements expressed in only a single tissue.

The genes expressed in a single tissue were further analyzed, and the results of the analyses are compiled in FIG. 7.

FIG. 7A is a matrix presenting the expression of all verified sequences that showed expression greater than 3 in at least one tissue. Each clone is represented by a column in the matrix. Each of the 10 tissues assayed is represented by a separate row in the matrix, and relative expression of a clone in that tissue is indicated at the respective node by intensity of green shading, with the intensity legend shown in panel B. The top row of the matrix ("EST Hit") contains "bioinformatic" rather than "physical" expression data – that is, presents the results returned by query of EST, NR and SwissProt databases using the probe sequence. The legend for "bioinformatic expression" (*i.e.*, degree of homology returned) is presented in panel C. Briefly, white is known, black is novel, with gray depicting nonidentical with significant homology (white: E values < 1e-100; gray: E values from 1e-

05 to 1e-99; black: E values > 1e-05).

As FIG. 7 readily shows, heart and brain were demonstrated to have the greatest numbers of genes that were shown to be uniquely expressed in the respective tissue. In brain, 200 uniquely expressed genes were identified; in heart, 150. The remaining tissues gave the following figures for uniquely expressed genes: liver, 100; lung, 70; fetal liver, 150; bone marrow, 75; placenta, 100; HeLa, 50; HBL, 100; and BT474, 50.

It was further observed that there were many more "novel" genes among those that were up-regulated in only one tissue, as compared with those that were down-regulated in only one tissue. In fact, it was found that ORFs whose expression was measurable in only a single of the tested tissues were represented in sequencing databases at a rate of only 11%, whereas 36% of the ORFs whose expression was measurable in 9 of the tissues were present in public databases. As for those ORFs expressed in all ten tissues, fully 45% were present in existing expressed sequence databases. These results are not unexpected, since genes expressed in a greater number of tissues have a higher likelihood of being, and thus of having been, discovered by EST approaches.

Comparison of Signal from Known and Unknown Genes

The normalized signal of the genes found to have high homology to genes present in the GenBank human EST database were compared to the normalized signal of those genes not found in the GenBank human EST database. The data are shown in FIG. 8.

FIG. 8 shows the normalized Cy3 signal intensity for all sequence-verified products with a BLAST Expect ("E") value of greater than 1e-30 (designated "unknown") upon query of existing EST, NR and SwissProt databases, and shows in blue the normalized Cy3 signal intensity for all

sequence-verified products with a BLAST Expect value of less than 1e-30 ("known"). Note that biological background noise has an averaged normalized Cy3 signal intensity of 0.2.

As expected, the most highly expressed of the ORFs were "known" genes. This is not surprising, since very high signal intensity correlates with very commonly-expressed genes, which have a higher likelihood of being found by EST sequence.

However, a significant point is that a large number of even the high expressers were "unknown". Since the genomic approach used to identify genes and to confirm their expression does not bias exons toward either the 3' or 5' end of a gene, many of these high expression genes will not have been detected in an end-sequenced cDNA library.

The significant point is that presence of the gene in an EST database is *not* a prerequisite for incorporation into a genome-derived microarray, and further, that arraying such "unknown" exons can help to assign function to as-yet undiscovered genes.

Verification of Gene Expression

To ascertain the validity of the approach described above to identify genes from raw genomic sequence, expression of two of the probes was assayed using reverse transcriptase polymerase chain reaction (RT PCR) and northern blot analysis.

Two microarray probes were selected on the basis of exon size, prior sequencing success, and tissue-specific gene expression patterns as measured by the microarray experiments. The primers originally used to amplify the two respective ORFs from genomic DNA were used in RT PCR against a panel of tissue-specific cDNAs (Rapid-Scan gene expression panel 24 human cDNAs) (OriGene Technologies,

Inc., Rockville, MD).

Sequence AL079300_1 was shown by microarray hybridization to be present in cardiac tissue, and sequence AL031734_1 was shown by microarray experiment to be present in placental tissue (data not shown). RT-PCR on these two sequences confirmed the tissue-specific gene expression as measured by microarrays, as ascertained by the presence of a correctly sized PCR product from the respective tissue type cDNAs.

Clearly, all microarray results cannot, and indeed should not, be confirmed by independent assay methods, or the high throughput, highly parallel advantages of microarray hybridization assays will be lost. However, in addition to the two RT-PCR results presented above, the observation that 1/3 of the arrayed genes exist in expression databases provides powerful confirmation of the power of our methodology – which combines bioinformatic prediction with expression confirmation using genome-derived single exon microarrays – to identify novel genes from raw genomic data.

To verify that the approach further provides correct characterization of the expression patterns of the identified genes, a detailed analysis was performed of the microarrayed sequences that showed high signal in brain.

For this latter analysis, sequences that showed high (normalized) signal in brain, but which showed very low (normalized) signal (less than 0.5, determined to be biological noise) in all other tissues, were further studied. There were 82 sequences that fit these criteria, approximately 2% of the arrayed elements. The 10 sequences showing the highest signal in brain in microarray hybridizations are detailed in Table 2, along with assigned function, if known or reasonably predicted.

Table 2

Function of the Most Highly Expressed Genes Expressed Only in Brain				
Microarray Sequence Name	Normalized Signal	Expression Ratio	Homology to EST present in GenBank	Gene Function as described by GenBank
AP000217-1	5.2	+7.7	High	S-100 protein, b-chain, Ca^{2+} binding protein expressed in central nervous system
AP000047-1	2.3		High	Unknown Function
AC006548-9	1.7		High	Similar to mouse membrane glyco-protein M6, expressed in central nervous system
AC007245-5	1.5		High	Similar to amphiphysin, a synaptic vesicle-associated protein. Ref 21
L44140-4	1.2	+2.0	High	Endothelial actin-binding protein found in nonmuscle filamin

AC004689-9	1.2	+3.5	High	Protein Phosphatase PP2A, neuronal/ downregulates activated protein kinases
AL031657-1	1.2	+3.0	High	Unknown function/ Contains the anhyrin motif, a common protein sequence motif
AC009266-2	1.1	+3.7	Low	Low homology to the Synaptotagmin I protein in rat/present at low levels throughout rat brain
AP000086-1	1.0	+2.7	Low	Unknown, very poor homology to collagen
AC004689-3	1.0		High	Protein Phosphatase PP2A, neuronal/ downregulates activated protein kinases

Of the ten sequences studied by these latter confirmatory approaches, eight were previously known. Of these eight, six had previously been reported to be important in the central nervous system or brain. The exon

giving the highest signal (AP00217-1) was found to be the gene encoding an S100B Ca^{2+} binding protein, reported in the literature to be highly and uniquely expressed in the central nervous system. Heizmann, *Neurochem. Res.* 9:1097 (1997).

A number of the brain-specific probe sequences (including AC006548-9, AC009266-2) did not have homology to any known human cDNAs in GenBank but did show homology to rat and mouse cDNAs. Sequences AC004689-9 and AC004689-3 were both found to be phosphatases present in neurons (Millward *et al.*, *Trends Biochem. Sci.* 24(5):186-191 (1999)). Two microarray sequences, AP000047-1 and AP000086-1 have unknown function, with AP000086-1 being absent from GenBank. Functionality can now be narrowed down to a role in the central nervous system for both of these genes, showing the power of designing microarrays in this fashion.

Next, the function of the chip sequences with the highest (normalized) signal intensity in brain, regardless of expression in other tissues, was assessed. In this latter analysis, we found expression of many more common genes, since the sequences were not limited to those expressed only in brain. For example, looking at the 20 highest signal intensity spots in brain, 4 were similar to tubulin (AC00807905; AF146191-2; AC007664-4; AF14191-2), 2 were similar to actin (AL035701-2; AL034402-1), and 6 were found to be homologous to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (AL035604-1; Z86090-1; AC006064-L, AC006064-K; AC035604-3; AC006064-L). These genes are often used as controls or housekeeping genes in microarray experiments of all types.

Other interesting genes highly expressed in brain were a ferritin heavy chain protein, which is reported in the literature to be found in brain and liver (Joshi *et al.*, *J. Neurol. Sci.* 134(Suppl):52-56 (1995)), a result

duplicated with the array. Other highly expressed chip sequences included a translation elongation factor 1□ (AC007564-4), a DEAD-box homolog (AL023804-4), and a Y-chromosome RNA-binding motif (Chai *et al.*, *Genomics* 49(2):283-89 (1998))(AC007320-3). A low homology analog (AP00123-1/2) to a gene, DSCR1, thought to be involved in trisomy 21 (Down's syndrome), showed high expression in both brain and heart, in agreement with the literature (Fuentes *et al.*, *Mol. Genet.* 4(10):1935-44 (1995)).

As a further validation of the approach, we selected the BAC AC006064 to be included on the array. This BAC was known to contain the GAPDH gene, and thus could be used as a control for the ORF selection process. The gene finding and exon selection algorithms resulted in choosing 25 exons from BAC AC006064 for spotting onto the array, of which four were drawn from the GAPDH gene. Table 3 shows the comparison of the average expression ratio for the 4 exons from BAC006064 compared with the average expression ratio for 5 different dilutions of a commercially available GAPDH cDNA (Clontech).

Table 3

Comparison of Expression Ratio, for each tissue, of GAPDH		
	AC006064 (n = 4)	Control (n = 5)
Bone Marrow	-1.81 ± 0.11	-1.85 ± 0.08
Brain	-1.41 ± 0.11	-1.17 ± 0.05
BT474	1.85 ± 0.09	1.66 ± 0.12
Fetal Liver	-1.62 ± 0.07	-1.41 ± 0.05
HBL100	1.32 ± 0.05	2.64 ± 0.12
Heart	1.16 ± 0.09	1.56 ± 0.10
HeLa	1.11 ±0.06	1.30 ± 0.15
Liver	-1.62 ± 0.22	-2.07 ±

Lung	-4.95 ± 0.93	-3.75 ± 0.21
Placenta	-3.56 ± 0.25	-3.52 ± 0.43

Each tissue shows excellent agreement between the experimentally chosen exons and the control, again demonstrating the validity of the present exon mining approach. In addition, the data also show the variability of expression of GAPDH within tissues, calling into question its classification as a housekeeping gene and utility as a housekeeping control in microarray experiments.

EXAMPLE 3

Representation of Sequence and Expression Data as a "Mondrian"

For each genomic clone processed for microarray as above-described, a plethora of information was accumulated, including full clone sequence, probe sequence within the clone, results of each of the three gene finding programs, EST information associated with the probe sequences, and microarray signal and expression for multiple tissues, challenging our ability to display the information.

Accordingly, we devised a new tool for visual display of the sequence with its attendant annotation which, in deference to its visual similarity to the paintings of Piet Mondrian, is hereinafter termed a "Mondrian". FIGS. 3 and 4 present the key to the information presented on a Mondrian.

FIG. 9 presents a Mondrian of BAC AC008172 (bases 25,000 to 130,000 shown), containing the carbamyl phosphate synthetase gene (AF154830.1). Purple background within the region shown as field 81 in FIG. 3 indicates all 37 known

exons for this gene.

As can be seen, GRAIL II successfully identified 27 of the known exons (73%), GENEFINDER successfully identified 37 of the known exons (100%), while DICTION identified 7 of the known exons (19%).

Seven of the predicted exons were selected for physical assay, of which 5 successfully amplified by PCR and were sequenced. These five exons were all found to be from the same gene, the carbamyl phosphate synthetase gene (AF154830.1).

The five exons were arrayed, and gene expression measured across 10 tissues. As is readily seen in the Mondrian, the five chip sequences on the array show identical expression patterns, elegantly demonstrating the reproducibility of the system.

FIG. 10 is a Mondrian of BAC AL049839. We selected 12 exons from this BAC, of which 10 successfully sequenced, which were found to form between 5 and 6 genes. Interestingly, 4 of the genes on this BAC are protease inhibitors. Again, these data elegantly show that exons selected from the same gene show the same expression patterns, depicted below the red line. From this figure, it is clear that our ability to find known genes is very good. A novel gene is also found from 86.6 kb to 88.6 kb, upon which all the exon finding programs agree. We are confident we have two exons from a single gene since they show the same expression patterns and the exons are proximal to each other. Backgrounds in the following colors indicate a known gene (top to bottom):
red = kallistatin protease inhibitor (P29622);
purple = plasma serine protease inhibitor (P05154);
turquoise = α1 anti-chymotrypsin (P01011); mauve = 40S ribosomal protein (P08865). Note that chip sequence 8 and 12 did not sequence verify.

EXAMPLE 4

Genome-Derived Single Exon Probes Useful For Measuring Human Gene Expression

The protocols set forth in Examples 1 and 2, *supra*, were applied to additional human genomic sequence as it became newly available in GenBank to identify unique exons in the human genome that could be shown to be expressed at significant levels in Fetal liver tissue.

These unique exons are within longer probe sequences. Each probe was completely sequenced on both strands prior to its use on a genome-derived single exon microarray; sequencing confirms the exact chemical structure of each probe. An added benefit of sequencing is that it placed us in possession of a set of single base-incremented fragments of the sequenced nucleic acid, starting from the sequencing primer 3' OH. (Since the single exon probes were first obtained by PCR amplification from genomic DNA, we were of course additionally in possession of an even larger set of single base incremented fragments of each of the 12,673 single exon probes, each fragment corresponding to an extension product from one of the two amplification primers.)

The structures of the 12,673 unique single exon probes are clearly presented in the Sequence Listing as SEQ ID Nos.: 1 - 12,673. The 16 nt 5' primer sequence and 16 nt 3' primer sequence present on the amplicon are not included in the sequence listing. The sequences of the exons present within each of these probes is presented in the Sequence Listing as SEQ ID Nos.: 12,674 - 25,129, respectively. It will be noted that some amplicons have more than one exon, some exons are contained in more than one amplicon.

As detailed in Example 2, expression was

demonstrated by disposing the amplicons as single exon probes on nucleic acid microarrays and then performing two-color fluorescent hybridization analysis; significant expression is based on a statistical confidence that the signal is significantly greater than negative biological control spots. The negative biological control is formed from spotted DNA sequences from a different species. Here, 32 sequences from E.Coli were spotted in duplicate to give a total of 64 spots.

For each hybridisation (each slide, each colour) the median value of the signal from all of the spots is determined. The normalised signal value is the arithmetic mean of the signal from duplicate spots divided by the population median.

Control spots are eliminated if there is more that a five-fold difference between each one of the duplicate spots raw signals.

The median of the signal from the remaining control spots is calculated and all subsequent calculations are done with normalised signals.

Control spots having a signal of greater than median + 2.4 (the value 2.4 is roughly 12 times the observed standard deviation of control spot populations) are eliminated. Spots with such high signals are considered to be "outliers".

The mean and standard deviation of the modified control spot populations are calculated.

The mean + 3x the standard deviation (mean + (3*SD)) is used as the signal threshold qualifier for that particular hybridisation. Thus, individual thresholds are determined for each channel and each hybridisation.

This means that, assuming that the data is distributed normally, there is a 99% confidence that any signal exceeding the threshold is significant.

The probes and their expression data are

presented in Table 4, set forth respectively in Example 5. Example 5 presents the subset of probes that is significantly expressed in the human Fetal liver and thus presents the subset of probes that was recognized to be useful for measuring expression of their cognate genes in human Fetal liver tissue.

The sequence of each of the exon probes identified by SEQ ID NOS.: 12,674 to 25,129 was individually used as a BLAST (or, for SWISSPROT, BLASTX) query to identify the most similar sequence in each of dbEST, SwissProt (BLASTX), and NR divisions of GenBank. Because the query sequences are themselves derived from genomic sequence in GenBank, only nongenomic hits from NR were scored.

The smallest in value of the BLAST (or BLASTX) expect ("E") scores for each query sequence across the three database divisions was used as a measure of the "expression novelty" of the probe's ORF. Table 4 is sorted in descending order based on this measure, reported as "Most Similar (top) Hit BLAST E Value". Those sequences for which no "Hit E Value" is listed are those exons which were found to have no similar sequences.

As sorted, Table 4 thus lists its respective probes (by "AMPLICON SEQ ID NO.:" and additionally by the SEQ ID NO:. of the exon contained within the probe:"EXON SEQ ID NO.:") from least similar to sequences known to be expressed (*i.e.*, highest BLAST E value), at the beginning of the table, to most similar to sequences known to be expressed (*i.e.*, lowest BLAST E value), at the bottom of the table.

Table 4 further provides, for each listed probe, the accession number of the database sequence that yielded the "Most Similar (top) Hit BLAST E Value", along with the name of the database in which the database sequence is found ("Top Hit Database Source").

Table 4 further provides SEQ ID NOS. corresponding to the predicted amino acid sequences where they have been determined for the probe and exon nucleotide sequences. These are set out as PEPTIDE SEQ ID NOS.:. The peptide sequences for a given exon are predicted as follows: Since each chip exon is a consensus sequence drawn from predictions from various exon finding programs (i.e. Grail, GeneFinder and GenScan), the multiple initial ORFs are first determined in a uniform way according to each prediction. In particular, the reading frame for predicting the first amino acid in the peptide sequence always starts with the first base of any codon and ends with the last base of non-termination codon. Next, for each strand of the exon, initial ORFs are merged into one or more final ORFs in an exhaustive process based on the following criteria: 1) the merging ORFs must be overlapping, and 2) the merging ORFs must be in the same frame.

The Sequence Listing, which is a superset of all of the data presented in Table 4, further includes, for each probe, the most similar hit, with accession number and BLAST E value, from the each of the three queried databases.

Table 4 further lists, for each probe, a portion of the descriptor for the top hit ("Top Hit Descriptor") as provided in the sequence database. For those ORFs that are similar in sequence, but nonidentical to known sequences (*e.g.*, those with BLAST E values between about 1e-05 and 1e-100), the descriptor reveals the likely function of the protein encoded by the probe's ORF.

Using BLAST E value cutoffs of 1e-05 (*i.e.*, 1×10^{-5}) and 1e-100 (*i.e.*, 1×10^{-100}) as evidence of similarity to sequences known to be expressed is of course arbitrary: in Example 2, *supra*, a BLAST E value of 1e-30 was used as the boundary when only two classes were to be defined for analysis (unknown, >1e-30; known <1e-30) (see also FIG. 8).

Furthermore, even when the "Most Similar (Top) Hit BLAST E Value" is low, *e.g.*, less than about 1e-100 – which is probative evidence that the query sequence has previously been shown to be expressed – the top hit is highly unlikely exactly to match the probe sequence.

First, such expression entries typically will not have the intronic and/or intergenic sequence present within the single exon probes listed in the Table. Second, even the ORF itself is unlikely in such cases to be present identically in the databases, since most of the EST and mRNA clones in existing databases include multiple exons, without any indication of the location of exon boundaries.

As noted, the data presented in Table 4 represent a proper subset of the data present within the attached sequence listing. For each amplicon probe (SEQ ID NOs.: 1 - 12,673) and probe exon (SEQ ID NOs.: 12,674 - 25,129, respectively), the sequence listing further provides, through iterated annotation fields <220> and <223>:

(a) the accession number of the BAC from which the sequence was derived ("MAP TO"), thus providing a link to the chromosomal map location and other information about the genomic milieu of the probe sequence;

(b) the most similar sequence provided by BLAST query of the EST database, with accession number and BLAST E value for the "hit";

(c) the most similar sequence provided by BLAST query of the GenBank NR database, with accession number and BLAST E value for the "hit"; and

(d) the most similar sequence provided by BLASTX query of the SWISSPROT database, with accession number and BLAST E value for the "hit".

EXAMPLE 5

Genome-Derived Single Exon Probes Useful For Measuring

Expression of Genes in Human Fetal liver

Table 4 (526 pages) presents expression, homology, and functional information for the genome-derived single exon probes that are expressed significantly in human Fetal liver.

CLAIMS

1. A spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human Fetal liver comprising a plurality single exon nucleic probes, said probes comprising any one of the nucleotide sequences set out in SEQ ID NOs: 1 - 12,673 or a complementary sequence, or a portion of such a sequence.

2. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably amplifiable.

3. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably isolatable from said plurality.

4. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 3 wherein said probes comprise any one of the nucleotide sequences set out in SEQ ID NOS.: 12,674 - 25,129.

5. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 4, wherein each of said plurality of probes is amplifiable using at least one common primer.

6. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 5 wherein the set comprises between 50 - 20,000 single exon nucleic acid probes.

7. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 6, wherein the

average length of the single exon nucleic acid probes is between 200 and 500 bp.

8. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 7, wherein at least 50% of said single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence.

9. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 8, wherein at least 50% of said single-exon nucleic acid probes lack homopolymeric stretches of A or T.

10. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 9 characterised in that said set of probes is addressably disposed upon a substrate.

11. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 10 wherein said substrate is selected from glass, amorphous silicon, crystalline silicon and plastic.

12. A microarray comprising a spatially addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 11.

13. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human Fetal liver comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 1 - 12,673 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid molecule expressed in the human Fetal liver.

14. A single exon nucleic acid probe as claimed in claim 13 comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 12,674 - 25,129 or a complementary sequence or a fragment thereof.

15. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human Fetal liver which is a nucleic acid molecule having a sequence encoding a peptide comprising a peptide sequence as set out in any of SEQ ID NOs.: 25,130 - 37,156, or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human Fetal liver.

16. A single exon nucleic acid probe as claimed in any one of claims 13 to 15 wherein said single exon nucleic acid probe comprises between 15 and 25 contiguous nucleotides of said SEQ ID NO.

17. A single exon nucleic acid probe as claimed in any one of claims 13 to 15, wherein said probe is between 3 - 25 kb in length.

18. A single exon nucleic acid probe as claimed in any one of claims 13 - 17, wherein said probe is DNA, RNA or PNA.

19. A single exon nucleic acid probe as claimed in any one of claims 13 - 18, wherein said probe is detectably labeled.

20. A single exon nucleic acid probe as claimed in any one of claims 13 - 19, wherein said probe lacks prokaryotic and bacteriophage vector sequence.

21. A single exon nucleic acid probe as claimed in any one

of claims 13 - 20, wherein said probe lacks homopolymeric stretches of A or T.

22. A method of measuring gene expression in a sample derived from human Fetal liver, comprising:

contacting the microarray of claim 12, with a first collection of detectably labeled nucleic acids, said first collection of nucleic acids derived from mRNA of human Fetal liver; and then

measuring the label detectably bound to each probe of said microarray.

23. A method of identifying exons in a eukaryotic genome, comprising:

algorithmically predicting at least one exon from genomic sequence of said eukaryote; and then

detecting specific hybridization of detectably labeled nucleic acids to a single exon probe,

wherein said detectably labeled nucleic acids are derived from mRNA from the Fetal liver of said eukaryote, said probe is a single exon probe having a fragment identical in sequence to, or complementary in sequence to, said predicted exon, said probe is included within a microarray according to claim 12, and said fragment is selectively hybridizable at high stringency.

24. A method of assigning exons to a single gene, comprising:

identifying a plurality of exons from genomic sequence according to the method of claim 23; and then

measuring the expression of each of said exons in a plurality of tissues and/or cell types using hybridization to single exon microarrays having a probe with said exon,

wherein a common pattern of expression of said exons in said plurality of tissues and/or cell types indicates that the exons should be assigned to a single gene.

25. A nucleic acid sequence as set out in any of SEQ ID NOs: 1 - 25,129 which encodes a peptide.

26. A peptide encoded by a sequence as set out in any of SEQ ID Nos: 1 - 25,129.

27. A peptide comprising a sequence as set out in any of SEQ ID Nos: 25,130 - 37,156.

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
475	13108	25800	4.41				
922	13535	26053	9.9				
1083	13688		2.9				
1345	13940	26462	10.32				
1656	14248	26782	2.59				
1678	14270	26803	5.03				
1763	14353	26899	1.73				
1785	14375	26919	0.99				
1792	14382	26927	9.24				
1935	14519	27075	1.21				
2021	14603	27168	3.24				
2210	14786	27360	4.38				
2318	14890	27465	2.04				
2607	15169	27735	0.89				
2607	15169	27736	0.89				
3220	15832	28311	1.65				
3496	16101	28576	1.22				
3566	16170	28652	10.28				
3617	16220		0.8				
3718	16319	28787	0.97				
4020	16618		0.94				
4275	16861	29310	1.53				
4348	16935	29376	8.4				
4368	16955	29396	0.74				
4368	16955	29397	0.74				
4430	17016		1.3				
4962	17537	29979	1.04				
5007	17580		0.59				
5054	17627	30071	0.61				
5197	17762	30187	5.95				
5212	17777	30198	1.32				
5462	18097	30415	2.1				
5462	18097	30416	2.1				
5615	18244		5.64				

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5783	18408		9.03				
5859	18244		4.85				
5910	18532	31257	0.84				
5915	18537	31262	3.16				
6173	24759	31552	1.41				
6200	18810	31579	1.65				
6548	19146		1.26				
6668	19264	32067	1				
6668	19264	32068	1				
7179	19711	32559	1.13				
7179	19711	32560	1.13				
7441	19965	32831	1.4				
7441	19965	32832	1.4				
8005	20547	33451	1.65				
8422	20962	33878	1.45				
8794	21333	34257	0.57				
8794	21333	34258	0.57				
9453	21979	34931	4.84				
9681	22180	35155	0.78				
9796	22294	35277	1.19				
9936	22431	35406	1.03				
10214	22709	35702	0.48				
10214	22709	35703	0.48				
10326	22820	35815	0.65				
10326	22820	35816	0.65				
10563	23099		3.06				
10725	24798	36268	2.46				
10906	23425		2.99				
11238	23769	36827	2.73				
11336	23034	36043	1.87				
11336	23034	36044	1.87				
11374	23826		2.59				
12117	24376		2.19				
12439	24576	30914	1.6				

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6203	18813	31583	14.37	9.9E+00	AJ239028.1	NT	Homo sapiens LSS gene, partial, exons 15, 16, 17 and 18
7948	20490	33400	1.66	9.8E+00	U32716.1	NT	Haemophilus influenzae Rd section 31 of 163 of the complete genome
9658	22157	35128	0.47	9.8E+00	Y18930.1	NT	Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2
9658	22157	35129	0.47	9.8E+00	Y18930.1	NT	Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2
7073	19645	32483	0.8	9.6E+00	AF065630.1	NT	Gallus gallus ornithine transcarbamylase (OTC) gene, exon 1
7073	19645	32484	0.8	9.6E+00	AF065630.1	NT	Gallus gallus ornithine transcarbamylase (OTC) gene, exon 1
10319	22813	35808	1.22	9.6E+00	AF242432.1	NT	Mus musculus Naip3 gene, exon 1; neuronal apoptosis inhibitory protein 1 (Naip1) and general transcription factor IIH polypeptide 2 (Gtf2h2) genes, complete cds
10319	22813	35809	1.22	9.6E+00	AF242432.1	NT	Mus musculus Naip3 gene, exon 1; neuronal apoptosis inhibitory protein 1 (Naip1) and general transcription factor IIH polypeptide 2 (Gtf2h2) genes, complete cds
2689	15247	27814	1.14	9.4E+00	L11433.1	NT	Dengue virus type 3 membrane protein (prM/M)/envelope glycoprotein (E) polyprotein mRNA, partial cds
2689	15247	27815	1.14	9.4E+00	L11433.1	NT	Dengue virus type 3 membrane protein (prM/M)/envelope glycoprotein (E) polyprotein mRNA, partial cds
2950	15568	28040	3.19	9.4E+00	AB043785.1	NT	Mus musculus AT3 gene for antithrombin, complete cds
8042	20584	33491	0.99	9.3E+00	AF130990.1	NT	Homo sapiens ectodysplasin-A receptor protein (EDAR) gene, exons 2, 3, and 4
8933	21471	34390	3.48	9.3E+00	P11210	SWISSPROT	IMMEDIATE-EARLY PROTEIN 1 (IE1) (IMMEDIATE-EARLY PHOSPHOPROTEIN PP89)
5500	18134	30543	2.82	9.1E+00	AF095609.1	NT	Leuciscus cephalus orientalis cytochrome b (cyt b) gene, partial cds; mitochondrial gene for mitochondrial product
5500	18134	30544	2.82	9.1E+00	AF095609.1	NT	Leuciscus cephalus orientalis cytochrome b (cyt b) gene, partial cds; mitochondrial gene for mitochondrial product
9351	21865		0.9	9.0E+00	P09241	SWISSPROT	RHODOPSIN
6186	18796	31584	5.12	8.9E+00	BE971806.1	EST_HUMAN	601651038R1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:3934592 3'
6517	19117	31907	1.9	8.7E+00	AB019788.1	NT	Cynops pyrrhogaster CpTbx3 premature mRNA, partial cds
6517	19117	31908	1.9	8.7E+00	AB019788.1	NT	Cynops pyrrhogaster CpTbx3 premature mRNA, partial cds
465	13099	25590	1.68	8.4E+00	5031804	NT	Homo sapiens insulin receptor substrate 1 (IRS1) mRNA
9376	20315	33217	3.8	8.1E+00	AJ131719.1	NT	Zea mays mRNA for legumain-like protease (see2a)
11048	23561		2.47	8.0E+00	P41820	SWISSPROT	BREFELDIN A RESISTANCE PROTEIN
8092	20633		0.76	7.6E+00	Z21489.1	NT	African swine fever virus NP1450L gene encoding RNA polymerase largest subunit
7384	19910		1.95	7.5E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
8302	20843	33764	1.54	7.5E+00	P35441	SWISSPROT	THROMBOSPONDIN 1 PRECURSOR
8302	20843	33765	1.54	7.5E+00	P35441	SWISSPROT	THROMBOSPONDIN 1 PRECURSOR
5968	18589	31324	3.35	7.4E+00	BF700517.1	EST_HUMAN	602128876F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285506 5'
8688	21227	34147	2.63	7.4E+00	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed In Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8688	21227	34148	2.63	7.4E+00	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
3006	15622	28099	3.19	7.2E+00	L12051.1	NT	Lycopersicon esculentum Mill. GTPase (SAR2) mRNA, complete cds
3006	15622	28100	3.19	7.2E+00	L12051.1	NT	Lycopersicon esculentum Mill. GTPase (SAR2) mRNA, complete cds
7097	19668	32507	0.7	7.2E+00	BE179090.1	EST_HUMAN	RC0-HT0613-200300-031-a07 HT0613 Homo sapiens cDNA
7203	19734	32585	1.22	7.1E+00	P28166	SWISSPROT	ZINC-FINGER PROTEIN 1 (ZINC-FINGER HOMEODOMAIN PROTEIN 1)
7203	19734	32586	1.22	7.1E+00	P28166	SWISSPROT	ZINC-FINGER PROTEIN 1 (ZINC-FINGER HOMEODOMAIN PROTEIN 1)
9516	22016		7.96	7.1E+00	AL161595.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 91
11263	23791	36848	3.2	7.1E+00	P05850	SWISSPROT	HYPOTHETICAL 17.3 KDA PROTEIN IN MRDA-PHPB INTERGENIC REGION
9892	22389	35367	3.35	7.0E+00	P48610	SWISSPROT	ARGININE KINASE (AK)
11129	23637	36679	1.87	7.0E+00	O22469	SWISSPROT	WD-40 REPEAT PROTEIN MSI3
8225	20766	33684	4.06	6.9E+00	P35679	SWISSPROT	60S RIBOSOMAL PROTEIN L4 (L2)
10253	22748	35736	1.2	6.9E+00	P44834	SWISSPROT	DNA MISMATCH REPAIR PROTEIN MUTS
7849	20391	33293	1.38	6.8E+00	W03412.1	EST_HUMAN	za07c11.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:291860 5'
7849	20391	33294	1.38	6.8E+00	W03412.1	EST_HUMAN	za07c11.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:291860 5'
9060	21597		1.13	6.8E+00	P36307	SWISSPROT	OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4) [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8]
10110	22605	35595	3.85	6.8E+00	Q03570	SWISSPROT	HYPOTHETICAL 157.0 KDA PROTEIN C38C10.5 IN CHROMOSOME III
5488	18122		0.69	6.6E+00	Q99028	SWISSPROT	CATECHOL-O-METHYLTRANSFERASE, SOLUBLE FORM (S-COMT)
9986	22481	35485	1.89	6.6E+00	Q9ZE07	SWISSPROT	URIDYLATE KINASE (UK) (URIDINE MONOPHOSPHATE KINASE) (UMP KINASE)
9986	22481	35486	1.89	6.6E+00	Q9ZE07	SWISSPROT	URIDYLATE KINASE (UK) (URIDINE MONOPHOSPHATE KINASE) (UMP KINASE)
11008	23522		2.13	6.6E+00	Q10309	SWISSPROT	PROBABLE CATION-TRANSPORTING ATPASE C8C3.05C
9108	21644	34584	7.21	6.5E+00	P03374	SWISSPROT	ENV POLYPROTEIN [CONTAINS: COAT PROTEIN GP52; COAT PROTEIN GP36]
10206	22701	35695	0.49	6.5E+00	BE866001.1	EST_HUMAN	601678435F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3960969 5'
9657	22156	35127	1.11	6.2E+00	AY010901.1	NT	Schizophyllum commune unknown mRNA
10455	22949	35958	0.53	6.2E+00	6754621	NT	Mus musculus mannosidase 2, alpha B1 (Man2b1), mRNA
7102	19672	32511	1.34	6.0E+00	BE780163.1	EST_HUMAN	601468031F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3871303 5'
9730	22228	35205	0.46	6.0E+00	AP000006.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 1166001-1485000 nt. position (6/7)
10407	22901	35896	0.6	6.0E+00	AE001862.1	NT	Deinococcus radiodurans R1 section 1 of 2 of the complete chromosome 2
10407	22901	35897	0.6	6.0E+00	AE001862.1	NT	Deinococcus radiodurans R1 section 1 of 2 of the complete chromosome 2
6643	19239	32042	6.67	5.9E+00	AF155142.1	NT	Mus musculus mixed lineage kinase 3 (Mlk3) and two pore domain K+ channel subunit (Kcnk6) genes, complete cds
3576	16180		1.18	5.8E+00	7661557	NT	Homo sapiens DESC1 protein (DESC1), mRNA
7215	19746	32601	0.67	5.7E+00	AF302046.1	NT	Mus musculus immunoglobulin scavenger receptor IgSR mRNA, complete cds
7215	19746	32602	0.67	5.7E+00	AF302046.1	NT	Mus musculus immunoglobulin scavenger receptor IgSR mRNA, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7570	20087		1.31	5.6E+00	P75080	SWISSPROT	DNA POLYMERASE III, ALPHA CHAIN POLC-TYPE (POLIII)
11349	23047	36059	2.66	5.6E+00	Q55276	SWISSPROT	LYCOPENE BETA CYCLASE
6399	19002	31780	0.73	5.5E+00	P47447	SWISSPROT	HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR HRCA
10658	23190		1.54	5.5E+00	AF175425.1	NT	Mus musculus DNA methyltransferase (Dnmt1) gene, exons 30, 31, and 32
11348	23046	36058	3.79	5.5E+00	P11990	SWISSPROT	PNEUMOLYSIN (THIOL-ACTIVATED CYTOLYSIN)
11578	24024		2.08	5.5E+00	AL161571.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 67
7009	19507	32326	1.2	5.4E+00	X02212.1	NT	Chicken alpha-cardiac actin gene
7009	19507	32327	1.2	5.4E+00	X02212.1	NT	Chicken alpha-cardiac actin gene
7369	19895		0.72	5.4E+00	Q99435	SWISSPROT	NEL PROTEIN PRECURSOR (NEL-RELATED PROTEIN 2)
7811	20354		1.58	5.4E+00	Q91082	SWISSPROT	VITELLOGENIN PRECURSOR (VTG) [CONTAINS: LIPOVITELLIN LV-1N; LIPOVITELLIN LV-1C; LIPOVITELLIN LV-2]
8734	21273	34193	0.78	5.4E+00	P40379	SWISSPROT	REP1 PROTEIN
8734	21273	34194	0.78	5.4E+00	P40379	SWISSPROT	REP1 PROTEIN
9949	22444	35423	1.33	5.4E+00	Q17094	SWISSPROT	RHODOPSIN
9949	22444	35424	1.33	5.4E+00	Q17094	SWISSPROT	RHODOPSIN
4899	17474	29930	1.52	5.3E+00	L43126.1	NT	Bovine immunodeficiency-like virus surface envelope gene, 5' end of cds
6614	19211		0.67	5.3E+00	P41779	SWISSPROT	HOMEOBOX PROTEIN CEH-20
8024	20568		3.71	5.3E+00	P54098	SWISSPROT	DNA POLYMERASE GAMMA (MITOCHONDRIAL DNA POLYMERASE CATALYTIC SUBUNIT)
8914	21452		0.62	5.3E+00	AB034990.1	NT	Homo sapiens HERPUD1 gene for stress protein Herp, complete cds
5655	18282		1.04	5.2E+00	BE184840.1	EST_HUMAN	QV4-HT0691-270400-186-f09 HT0691 Homo sapiens cDNA
10274	22769		0.78	5.2E+00	AF248070.1	NT	Drosophila orientacea R1B retrotransposable element reverse transcriptase gene, partial cds
11074	23586		2.1	5.2E+00	Q10136	SWISSPROT	HYPOTHETICAL 61.1 KD PROTEIN C23E2.03C IN CHROMOSOME I
8892	21430	34354	0.88	5.1E+00	O16005	SWISSPROT	RHODOPSIN
9739	22237	35217	0.97	5.1E+00	P09182	SWISSPROT	COLICIN N IMMUNITY PROTEIN (MICROCIN N IMMUNITY PROTEIN)
6430	19033	31817	0.85	5.0E+00	BF310443.1	EST_HUMAN	601894910F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124114 5'
10096	22591		0.69	5.0E+00	BF308561.1	EST_HUMAN	601890420F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131509 5'
10327	22821	35817	3.37	5.0E+00	AF162445.2	NT	Canis familiaris skeletal muscle chloride channel ClC-1 (CLCN1) mRNA, complete cds
11170	23677	36723	13.54	5.0E+00	Z83860.1	NT	Mycobacterium tuberculosis H37Rv complete genome; segment 103/162
10131	22626		0.71	4.9E+00	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
4135	16727		12.06	4.8E+00	AF185255.1	NT	Eunice australis histone H3 (H3) gene, partial cds
8095	20838	33547	0.65	4.8E+00	BF367909.1	EST_HUMAN	RC3-GN0042-100800-011-c10 GN0042 Homo sapiens cDNA
8478	21017		4.95	4.8E+00	AW750067.1	EST_HUMAN	PM0-BT0547-310100-002-b04 BT0547 Homo sapiens cDNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
311	12966	25454	1.88	4.7E+00	BF240552.1	EST_HUMAN	601875654F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4099716 5'
312	12966	25454	1.68	4.7E+00	BF240552.1	EST_HUMAN	601875654F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4099716 5'
3312	15923	28399	1.08	4.7E+00	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
9124	21659	34601	1.09	4.6E+00	BE646437.1	EST_HUMAN	7e86g10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3292098 3' similar to TR:O75140 O75140 KIAA0645 PROTEIN. ;contains element PTR5 repetitive element ;
9124	21659	34602	1.09	4.6E+00	BE646437.1	EST_HUMAN	7e86g10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3292098 3' similar to TR:O75140 O75140 KIAA0645 PROTEIN. ;contains element PTR5 repetitive element ;
10290	22785		0.77	4.6E+00	AF240786.1	NT	Homo sapiens glutathione S-transferase theta 2 (GSTT2) and glutathione S-transferase theta 1 (GSTT1) genes, complete cds
11474	23924	36994	1.99	4.5E+00	AE001044.1	NT	Archaeoglobus fulgidus section 63 of 172 of the complete genome
11596	24039	37108	1.67	4.5E+00	BF668841.1	EST_HUMAN	602123238F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4280216 5'
3076	15691	28164	1.53	4.4E+00	BF530893.1	EST_HUMAN	602072585F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215284 5'
3076	15691	28165	1.53	4.4E+00	BF530893.1	EST_HUMAN	602072585F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215284 5'
6349	18954		1.8	4.4E+00	X13414.1	NT	Murine I gene for MHC class II(Ia) associated invariant chain
6266	18874		0.82	4.3E+00	AF059679.1	NT	Homo sapiens neutrophil collagenase (CLGNA) gene, promoter region and 5'UTR
7464	19986	32851	2.36	4.3E+00	Y13402.1	NT	Plasmodium falciparum R29R+var1 gene, exon 1
7611	20124	33001	0.84	4.3E+00	AE001222.1	NT	Treponema pallidum section 38 of 87 of the complete genome
10741	23266	36282	8.92	4.3E+00	AF240786.1	NT	Homo sapiens glutathione S-transferase theta 2 (GSTT2) and glutathione S-transferase theta 1 (GSTT1) genes, complete cds
5708	18334		3.21	4.2E+00	P16444	SWISSPROT	MICROSOMAL DIPEPTIDASE PRECURSOR (MDP) (DEHYDROPEPTIDASE-I) (RENAL DIPEPTIDASE) (RDP)
5780	18405	31121	1.46	4.2E+00	P51826	SWISSPROT	LAF-4 PROTEIN (LYMPHOID NUCLEAR PROTEIN)
6869	19603	32435	1.86	4.2E+00	P13983	SWISSPROT	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)
6869	19603	32436	1.86	4.2E+00	P13983	SWISSPROT	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)
8890	21428	34353	4.95	4.2E+00	AI809013.1	EST_HUMAN	wf67g03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2360692 3'
9832	22330	35312	2.07	4.2E+00	P31368	SWISSPROT	NUBBIN PROTEIN (TWAIN PROTEIN) (POU DOMAIN PROTEIN 1) (PDM-1) (DPOU-19) (DOCT1)
7166	19698	32545	0.81	4.1E+00	BE253668.1	EST_HUMAN	601110727F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3351534 5'
7264	19792	32648	1.7	4.1E+00	BF247939.1	EST_HUMAN	601859030F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4069758 5'
7657	20169	33056	8.1	4.1E+00	O23810	SWISSPROT	YY1 PROTEIN PRECURSOR
7759	20267	33163	4.03	4.1E+00	P28964	SWISSPROT	GENE 68 PROTEIN
7759	20267	33164	4.03	4.1E+00	P28964	SWISSPROT	GENE 68 PROTEIN
7857	20399	33306	2.78	4.1E+00	U57503.1	NT	Pan troglodytes novel repetitive solo LTR element in the RNU2 locus
9459	21985	34939	0.63	4.1E+00	P11253	SWISSPROT	50S RIBOSOMAL PROTEIN L4
9590	22090	35054	2.26	4.1E+00	BF692425.1	EST_HUMAN	602247938F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4333209 5'

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10208	22703		0.5	4.1E+00	P46414	SWISSPROT	CYCLIN-DEPENDENT KINASE INHIBITOR 1B (CYCLIN-DEPENDENT KINASE INHIBITOR P27) (P27KIP1)
10499	22993	36003	0.62	4.1E+00	O84242	SWISSPROT	3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE III (BETA-KETOACYL-ACP SYNTHASE III) (KAS III)
10765	23289		2.97	4.1E+00	P09716	SWISSPROT	HYPOTHETICAL PROTEIN HVLF1
10851	23372		13.84	4.1E+00	BE885880.1	EST_HUMAN	601507510F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909051 5'
3599	16203		0.82	4.0E+00	P38229	SWISSPROT	GLC7-INTERACTING PROTEIN 1
5650	19515	32336	0.74	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
5650	19515	32337	0.74	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
7017	19515	32336	0.95	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
7017	19515	32337	0.95	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
7240	19769	32625	1.34	4.0E+00	O33010	SWISSPROT	CELL DIVISION PROTEIN FTSY HOMOLOG
10070	22565	35560	0.6	4.0E+00	AE002132.1	NT	Ureaplasma urealyticum section 33 of 59 of the complete genome
10158	22653	35647	0.49	4.0E+00	Q00511	SWISSPROT	URICASE (URATE OXIDASE)
10158	22653	35648	0.49	4.0E+00	Q00511	SWISSPROT	URICASE (URATE OXIDASE)
11423	23874	36937	3.99	4.0E+00	P07564	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
11423	23874	36938	3.99	4.0E+00	P07564	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
3550	16154	28636	4.79	3.9E+00	X64518.1	NT	N.tabacum chitinase gene 50 for class I chitinase C
4413	16998		0.74	3.9E+00	AF055466.1	NT	Mus musculus seminal vesicle secretory protein 99 (MSVSP99) gene, promoter region
5839	18463	31186	3.08	3.9E+00	BE814357.1	EST_HUMAN	MR0-BN0070-300500-028-h05 BN0070 Homo sapiens cDNA
5839	18463	31187	3.08	3.9E+00	BE814357.1	EST_HUMAN	MR0-BN0070-300500-028-h05 BN0070 Homo sapiens cDNA
6746	19339	32145	0.71	3.9E+00	AF298209.1	NT	Dictyostelium discoideum non-LTR retrotransposon TRE5-B, polyprotein (gag) and group-specific antigen (pol) genes, complete cds
6792	19383	32198	0.72	3.9E+00	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
6955	19532	32357	4.12	3.9E+00	P39299	SWISSPROT	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN AIDB-RPSF INTERGENIC REGION
7398	19923	32787	6.09	3.9E+00	M23907.1	NT	Human MHC class II lymphocyte antigen (DPw4-beta-1) gene, exon 2

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8259	20800	33718	2.15	3.9E+00	X65865.1	NT	X.laevis mRNA for M4 muscarinic receptor
11269	23007	36014	3.27	3.9E+00	Y18000.1	NT	Homo sapiens NF2 gene
11291	23743	36800	1.62	3.9E+00	AA661489.1	EST_HUMAN	nr18a12.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:1168318 similar to gb:A10416 METALLOPROTEINASE INHIBITOR 1 PRECURSOR (HUMAN);
2658	15217		1.1	3.8E+00	AE001562.1	NT	Helicobacter pylori, strain J99 section 123 of 132 of the complete genome
6525	19125	31918	0.78	3.8E+00	Q57830	SWISSPROT	HYPOTHETICAL PROTEIN MJ0385
8371	20911	33831	1.06	3.8E+00	D44725.1	EST_HUMAN	HUMSUPY135 Human brain cDNA Homo sapiens cDNA clone 148
9710	22208		0.55	3.8E+00	AJ390961.1	NT	Streptococcus oralis partial xpt gene for xanthine phosphoribosyltransferase, strain NCTC7864
4092	16687	29144	13.56	3.7E+00	AL161539.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 39
7218	19749		0.79	3.7E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
8642	21181		0.53	3.7E+00	4503950	NT	Homo sapiens glucokinase (hexokinase 4, maturity onset diabetes of the young 2) (GCK), nuclear gene encoding mitochondrial protein, mRNA
9105	21641	34581	0.68	3.7E+00	U43541.1	NT	Mus musculus laminin beta 2 gene, exons 17-33, and complete cds
11308	23801	36861	3.11	3.7E+00	BF669279.1	EST_HUMAN	602120551F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4277748 5'
11308	23801	36862	3.11	3.7E+00	BF669279.1	EST_HUMAN	602120551F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4277748 5'
11767	24158		1.28	3.7E+00	AB013746.3	NT	Gallus gallus mRNA for hypoxia-inducible factor-1 alpha, complete cds
619	13246	25719	2.6	3.6E+00	AV761055.1	EST_HUMAN	AV761055 MDS Homo sapiens cDNA clone MDSBUE10 5'
4917	17492		0.99	3.6E+00	AL161472.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2
8487	21026	33942	0.76	3.6E+00	D12367.1	EST_HUMAN	HUM000TB08 Liver HepG2 cell line. Homo sapiens cDNA clone tb08
8487	21026	33943	0.76	3.6E+00	D12367.1	EST_HUMAN	HUM000TB08 Liver HepG2 cell line. Homo sapiens cDNA clone tb08
8579	21118	34038	4.02	3.6E+00	AE004447.1	NT	Pseudomonas aeruginosa PA01, section 8 of 529 of the complete genome
8579	21118	34039	4.02	3.6E+00	AE004447.1	NT	Pseudomonas aeruginosa PA01, section 8 of 529 of the complete genome
10733	23259		4.32	3.6E+00	M96795.1	NT	Escherichia coli glycerophosphate dehydrogenase (glpD) gene, partial cds; and the translation start site has been verified (glpE), the translation start site has been verified (glpG), and repressor protein (glpR) genes, complete cds
3284	15895	28373	1.08	3.5E+00	AF221538.1	NT	Cryptosporidium felis heat shock protein 70 (HSP70) gene, partial cds
6151	18764		1.06	3.5E+00	L42898.1	NT	Borrelia burgdorferi (strain 25015) outer surface protein (ospC) gene, partial cds
6360	18964	31742	0.92	3.5E+00	R19745.1	EST_HUMAN	yg40c08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:34940 5'
8421	20961		0.55	3.5E+00	P24557	SWISSPROT	THROMBOXANE-A SYNTHASE (TXA SYNTHASE) (TXS)
8962	21500	34421	0.88	3.5E+00	AA190998.1	EST_HUMAN	zp86b04.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:627055 3' similar to contains Alu repetitive element;contains element MSR1 repetitive element ;
8962	21500	34422	0.88	3.5E+00	AA190998.1	EST_HUMAN	zp86b04.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:627055 3' similar to contains Alu repetitive element;contains element MSR1 repetitive element ;
9414	21923	34872	1.12	3.5E+00	AL161553.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 53

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1559	14151	26883	4.49	3.4E+00	AF254577.1	NT	Brassica napus RPB5d mRNA, complete cds
2612	15174	27742	1.02	3.4E+00	AL163278.2	NT	Homo sapiens chromosome 21 segment HS21C078
7397	19922	32786	2.85	3.4E+00	P04052	SWISSPROT	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT
7690	20199	33086	0.86	3.4E+00	P04052	SWISSPROT	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT
8611	21150		0.69	3.4E+00	U65408.1	NT	Human alternatively spliced potassium channels ROM-K1, ROM-K2, ROM-K3, ROM-K4, ROM-K5, and ROM-K6 (KCNJ1) gene, complete cds
9003	21540	34470	0.7	3.4E+00	AJ229042.1	NT	Homo sapiens 959 kb contig between AML1 and CBR1 on chromosome 21q22, segment 2/3
9040	21577	34508	0.5	3.4E+00	AJ250567.1	NT	Homo sapiens partial TM4SF2 gene for tetraspanin protein, exon 6
10165	22660	35655	3.61	3.4E+00	AF013167.1	NT	Saccharomyces cerevisiae MSS1 gene, complete cds
11406	23857	36922	1.98	3.4E+00	L77570.1	NT	Homo sapiens DiGeorge syndrome critical region, centromeric end
6218	18828	31601	0.9	3.3E+00	Q09669	SWISSPROT	PUTATIVE IRON ALCOHOL DEHYDROGENASE
6218	18828	31602	0.9	3.3E+00	Q09669	SWISSPROT	PUTATIVE IRON ALCOHOL DEHYDROGENASE
7834	20376	33281	0.88	3.3E+00	AF111168.2	NT	Homo sapiens serine palmitoyl transferase, subunit II gene, complete cds; and unknown genes
10361	22855	35847	0.87	3.3E+00	AP001511.1	NT	Bacillus halodurans genomic DNA, section 5/14
10361	22855	35848	0.87	3.3E+00	AP001511.1	NT	Bacillus halodurans genomic DNA, section 5/14
526	13158	25640	1.72	3.2E+00	X96422.1	NT	D.rerio zp-50 POU gene
4098	13158	25640	0.7	3.2E+00	X96422.1	NT	D.rerio zp-50 POU gene
4835	17413	29866	1.24	3.2E+00	4502404	NT	Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), mRNA
5757	18383	31095	1.34	3.2E+00	P54924	SWISSPROT	SQUALENE--HOPENE CYCLASE
5757	18383	31096	1.34	3.2E+00	P54924	SWISSPROT	SQUALENE--HOPENE CYCLASE
5787	18412	31128	2.45	3.2E+00	P12783	SWISSPROT	PHOSPHOGLYCERATE KINASE, CYTOSOLIC
5787	18412	31129	2.45	3.2E+00	P12783	SWISSPROT	PHOSPHOGLYCERATE KINASE, CYTOSOLIC
6448	19049	31834	1.86	3.2E+00	P18931	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4
6448	19049	31835	1.86	3.2E+00	P18931	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4
7601	20114	32991	0.84	3.2E+00	P04275	SWISSPROT	VON WILLEBRAND FACTOR PRECURSOR (VWF)
7750	20258	33154	2.22	3.2E+00	Y13655.1	NT	Chlamydomonas reinhardtii chloroplast DNA for rps9, ycf4, ycf3, rps18 genes
7750	20258	33155	2.22	3.2E+00	Y13655.1	NT	Chlamydomonas reinhardtii chloroplast DNA for rps9, ycf4, ycf3, rps18 genes
8960	21498		4.84	3.2E+00	P13061	SWISSPROT	PERIPLASMIC [NIFE] HYDROGENASE SMALL SUBUNIT (NIFE HYDROGENLYASE SMALL CHAIN)
9449	21975	34927	1.52	3.2E+00	M36383.1	NT	S.cerevisiae threonine deaminase (ILV1) gene, complete cds
10047	22542	35539	1.91	3.2E+00	AB016081.2	NT	Oryzias latipes OIGC6 gene for guanylyl cyclase C, complete cds
11727	24133		4.08	3.2E+00	L33836.1	NT	Sus scrofa choline acetyltransferase gene, promoter region
6035	18654	31396	2.24	3.1E+00	Q10135	SWISSPROT	HYPOTHETICAL 142.5 KD PROTEIN C23E2.02 IN CHROMOSOME I
7421	19945	32810	0.97	3.1E+00	P52178	SWISSPROT	TRIOSE PHOSPHATE/PHOSPHATE TRANSLOCATOR, NON-GREEN PLASTID PRECURSOR (CTPT)

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7711	20220		1.09	3.1E+00	AF303225.1	NT	Bacillus alcalophilus pectate lyase (pelE) gene, complete cds
8538	21077	33995	4.27	3.1E+00	P49894	SWISSPROT	TYPE I IODOTHYRONINE DEIODINASE (TYPE-I 5'DEIODINASE) (DIOI) (TYPE 1 DI) (5DI)
8538	21077	33996	4.27	3.1E+00	P49894	SWISSPROT	TYPE I IODOTHYRONINE DEIODINASE (TYPE-I 5'DEIODINASE) (DIOI) (TYPE 1 DI) (5DI)
9183	21760		3.77	3.1E+00	Q14957	SWISSPROT	GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C)
9249	21775	34726	0.52	3.1E+00	Q01149	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
9810	22308	35292	0.75	3.1E+00	7524759	NT	Chlorella vulgaris chloroplast, complete genome
9899	22396		0.56	3.1E+00	Q10125	SWISSPROT	HYPOTHETICAL 56.3 KD PROTEIN F52C9.5 IN CHROMOSOME III
10239	22734	35726	4.7	3.1E+00	P49365	SWISSPROT	DEOXYHYPUSINE SYNTHASE (DHS)
11338	23036		2.91	3.1E+00	P33515	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
11355	23809		7.48	3.1E+00	S56660.1	NT	retinoic acid nuclear receptor isoform beta 2 [mice, embryonal carcinoma cell line, PCC7-MZ1, mRNA, 2971 nt]
12490	24819		1.38	3.1E+00	U77666.1	NT	Brassica rapa pollen coat protein homolog (BAN103) gene, complete cds
5541	18173	30588	1.68	3.0E+00	X53096.1	NT	S.aureus genes encoding Sau96I DNA methyltransferase and Sau96I restriction endonuclease
6673	19269	32073	0.72	3.0E+00	X56037.1	NT	Corynebacterium glutamicum thrC gene for threonine synthase (EC 4.2.99.2)
6673	19269	32074	0.72	3.0E+00	X56037.1	NT	Corynebacterium glutamicum thrC gene for threonine synthase (EC 4.2.99.2)
7209	19740		10.44	3.0E+00	P18406	SWISSPROT	CYR61 PROTEIN PRECURSOR (3CH61)
7247	19776		0.77	3.0E+00	Q13201	SWISSPROT	ENDOTHELIAL CELL MULTIMERIN PRECURSOR
8838	21377		1.33	3.0E+00	X67838.1	NT	B.napus DNA for myrosinase
10195	22690	35683	0.53	3.0E+00	Q58605	SWISSPROT	S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINE ADENOSYLTRANSFERASE) (ADOMET SYNTHETASE)
10527	23064	36075	1.62	3.0E+00	Q16181	SWISSPROT	CDC10 PROTEIN HOMOLOG
10888	23409	36426	7.04	3.0E+00	P51842	SWISSPROT	RETINAL GUANYLYL CYCLASE 2 PRECURSOR (GUANYLATE CYCLASE 2F, RETINAL) (RETGC-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F)
10888	23409	36427	7.04	3.0E+00	P51842	SWISSPROT	RETINAL GUANYLYL CYCLASE 2 PRECURSOR (GUANYLATE CYCLASE 2F, RETINAL) (RETGC-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F)
2055	14636	27207	2.32	2.9E+00	AE002225.2	NT	Chlamydophila pneumoniae AR39, section 53 of 94 of the complete genome
6224	18833		0.68	2.9E+00	AB026033.1	NT	Bonapartia pedaliota mitochondrial DNA for 16S ribosomal RNA
6989	19487	32309	3.74	2.9E+00	Z36879.1	NT	F.pringlei gdcsPA gene for P-protein of the glycine cleavage system
7262	19790	32644	4.37	2.9E+00	O14514	SWISSPROT	BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1 PRECURSOR

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7262	19790	32645	4.37	2.9E+00	O14514	SWISSPROT	BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1 PRECURSOR
7479	20001	32868	6.04	2.9E+00	P46589	SWISSPROT	ADHERENCE FACTOR (ADHESION AND AGGREGATION MEDIATING SURFACE ANTIGEN)
7809	20352	33260	0.67	2.9E+00	P05844	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: MAJOR STRUCTURAL PROTEIN VP2; NONSTRUCTURAL PROTEIN VP4; MINOR STRUCTURAL PROTEIN VP3]
7809	20352	33261	0.67	2.9E+00	P05844	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: MAJOR STRUCTURAL PROTEIN VP2; NONSTRUCTURAL PROTEIN VP4; MINOR STRUCTURAL PROTEIN VP3]
8041	20583	33490	0.89	2.9E+00	BF344171.1	EST_HUMAN	602017413F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4153059 5'
1504	14096	26634	4.87	2.8E+00	AF186398.1	NT	Buxus harlandii maturase K (matK) gene, partial cds; chloroplast gene for chloroplast product
1675	14267		3.45	2.8E+00	AL161552.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 52
7348	19874	32740	4.88	2.8E+00	8393724	NT	Mus musculus endomucin (LOC53423), mRNA
9531	22031		0.57	2.8E+00	BE565182.1	EST_HUMAN	601342758F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3684807 5'
10569	19874	32740	1.68	2.8E+00	8393724	NT	Mus musculus endomucin (LOC53423), mRNA
251	12911	25394	9.31	2.7E+00	6679306	NT	Mus musculus per-hexamer repeat gene 3 (Phxr3), mRNA
251	12911	25395	9.31	2.7E+00	6679306	NT	Mus musculus per-hexamer repeat gene 3 (Phxr3), mRNA
5740	18366	31073	1.2	2.7E+00	L14005.1	NT	Homo sapiens apoA polymorphism Kringle IV gene, exons 1 and 2
8088	20629		0.8	2.7E+00	U15947.1	NT	Ipomoea purpurea chalcone synthase (CHSB) gene including complete 5'UTR and complete cds
8898	21436		1.68	2.7E+00	AL116459.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
9353	20292	33191	0.63	2.7E+00	AW088191.1	EST_HUMAN	xc88e12.x1 NCI_CGAP_Brn35 Homo sapiens cDNA clone IMAGE:2591374 3' similar to gb:M17733 THYMOSIN BETA-4 (HUMAN);
10394	22888		1.48	2.7E+00	BE063527.1	EST_HUMAN	CM0-BT0281-031199-087-h04 BT0281 Homo sapiens cDNA
4781	17362	29812	4.97	2.6E+00	AF068749.1	NT	Mus musculus sphingosine kinase (SPHK1b) mRNA, complete cds
5736	18362	31068	1.94	2.6E+00	6755601	NT	Mus musculus SRY-box containing gene 13 (Sox13), mRNA
5736	18362	31069	1.94	2.6E+00	6755601	NT	Mus musculus SRY-box containing gene 13 (Sox13), mRNA
5992	18612		2.42	2.6E+00	Y17062.1	NT	Mycobacterium fortuitum furA II gene
7689	20198		5.98	2.6E+00	AF235502.1	NT	Mus musculus SH2-containing inositol 5-phosphatase (Ship) gene, exons 16 through 27, and complete cds
8003	20545	33447	1.08	2.6E+00	AJ132180.1	NT	faba bean necrotic yellows virus C2-Eg gene, isolate Egyptian EV1-93
8003	20545	33448	1.08	2.6E+00	AJ132180.1	NT	faba bean necrotic yellows virus C2-Eg gene, isolate Egyptian EV1-93
9576	22076	35039	3.02	2.6E+00	AL161540.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40
10257	22752		1.51	2.6E+00	9055193	NT	Mus musculus cleavage and polyadenylation specificity factor 3 (Cpsf3), mRNA
10907	23426	36443	1.69	2.6E+00	AF143675.1	NT	Hantavirus Z10 segment M G1/G2 glycoprotein (Z10) gene, complete cds
12390	24986		2.78	2.6E+00	11419220	NT	Homo sapiens ATP-binding cassette, sub-family B (MDR/TAP), member 4 (ABCB4), mRNA
1513	14105	26640	2.29	2.5E+00	AJ271844.1	NT	Aspergillus nidulans recQ gene for DNA helicase, exons 1-4
1513	14105	26641	2.29	2.5E+00	AJ271844.1	NT	Aspergillus nidulans recQ gene for DNA helicase, exons 1-4

WO 01/57277

PCT/US01/00669

11/526

Table 4

Single Exon Probes Expressed In Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5981	18601	31334	1.71	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
5981	18601	31335	1.71	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6586	18601	31334	1.39	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6586	18601	31335	1.39	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6828	19418	32234	0.73	2.5E+00	D30052.1	NT	Vibrio cholerae cbxA gene and cbxB gene for cholera toxins, complete cds
7736	20244	33135	1.05	2.5E+00	AW949158.1	EST_HUMAN	QV4-FT0005-110500-205-g07 FT0005 Homo sapiens cDNA
9032	21569	34498	1.75	2.5E+00	D50307.1	NT	Rice DNA for aldolase C-1, complete cds
9766	22264	35247	0.88	2.5E+00	BE297758.1	EST_HUMAN	601175779F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3531090 5'
11724	24131		1.66	2.5E+00	AF289665.1	NT	Mus musculus EIF4H gene, partial cds; LIMK1 gene, complete cds; and ELN gene, partial cds
3047	15663	28144	0.9	2.4E+00	M24282.1	NT	Chicken alpha-3 collagen type VI mRNA, 3' end
5033	17607	30052	6.76	2.4E+00	4503352	NT	Homo sapiens double C2-like domains, alpha (DOC2A) mRNA
6161	18774	31536	4.02	2.4E+00	P02843	SWISSPROT	VITELLOGENIN I PRECURSOR (YOLK PROTEIN 1)
8082	20624	33536	1.99	2.4E+00	P26842	SWISSPROT	CD27L RECEPTOR PRECURSOR (T-CELL ACTIVATION ANTIGEN CD27) (T14)
8082	20624	33537	1.99	2.4E+00	P26842	SWISSPROT	CD27L RECEPTOR PRECURSOR (T-CELL ACTIVATION ANTIGEN CD27) (T14)
8153	20694		2.33	2.4E+00	AE001486.1	NT	Helicobacter pylori, strain J99 section 47 of 132 of the complete genome
8585	21124		1.62	2.4E+00	AW875126.1	EST_HUMAN	RC2-PT0004-031299-011-d05 PT0004 Homo sapiens cDNA
8762	21301	34222	8.16	2.4E+00	P24091	SWISSPROT	ENDOCHITINASE B PRECURSOR (CHN-B)
9951	22446	35427	2.59	2.4E+00	P13673	SWISSPROT	SKIN GRANULE PROTEIN PRECURSOR
9951	22446	35428	2.59	2.4E+00	P13673	SWISSPROT	SKIN GRANULE PROTEIN PRECURSOR
10017	22512	35504	1.62	2.4E+00	X92511.1	NT	H.sapiens CTGF gene and promoter region
10141	22636		7.38	2.4E+00	P09099	SWISSPROT	XYLULOSE KINASE (XYLULOKINASE)
10225	22720	35710	1.63	2.4E+00	BE326702.1	EST_HUMAN	hr63f06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3133187 3'
10225	22720	35711	1.63	2.4E+00	BE326702.1	EST_HUMAN	hr63f06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3133187 3'
10483	22977	35986	1.27	2.4E+00	Q51481	SWISSPROT	DENITRIFICATION REGULATORY PROTEIN NIRQ
10958	23473	36498	1.69	2.4E+00	Y14079.1	NT	Bacillus subtilis chromosomal DNA, region 75 degrees: glpPFKD operon and downstream
11237	23768	36826	2.27	2.4E+00	AF156652.2	NT	Fragaria x ananassa cytosolic ascorbate peroxidase (ApxSC) gene, ApxSC-c allele, complete cds
1296	13890	26413	11.15	2.3E+00	Z46724.1	NT	G.domesticus artificial single chain antibody gene (L3)
4199	16788		1.65	2.3E+00	AJ401081.1	NT	Bos taurus partial cytb gene for cytochrome b
6000	18620		0.91	2.3E+00	N86245.1	EST_HUMAN	J7340F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone J7340 5' similar to PROLYLCARBOXYPEPTIDASE
7477	19999	32884	2.54	2.3E+00	6978554	NT	Rattus norvegicus ATPase, Ca++ transporting, ubiquitous (Atp2a3), mRNA
7593	25120		4.61	2.3E+00	P07199	SWISSPROT	MAJOR CENTROMERE AUTOANTIGEN B (CENTROMERE PROTEIN B) (CENP-B)
7756	20264	33159	1.06	2.3E+00	X60265.1	NT	M.mazei dnaK and dnaJ genes homologues coding for DnaK and DnaJ

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Accession No.	Top Hit Database Source	Top Hit Descriptor
9038	21575	34505	0.53	2.3E+00	5835317	NT	Polypterus ornatipinnis mitochondrion, complete genome
9097	21633	34572	1.79	2.3E+00	Q11127	SWISSPROT	ALPHA-(1,3)-FUCOSYLTRANSFERASE (GALACTOSIDE 3-L-FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 4) (FUCT-IV)
10681	23213	36224	2	2.3E+00	Q07076	SWISSPROT	ANNEXIN VII (SYNEXIN)
11612	24055	37119	2.92	2.3E+00	BF541987.1	EST_HUMAN	602069121F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4068173 5'
11612	24055	37120	2.92	2.3E+00	BF541987.1	EST_HUMAN	602069121F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4068173 5'
11950	24278	31020	7.31	2.3E+00	BE895237.1	EST_HUMAN	601433673F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3918643 5'
4089	16685	29143	91.07	2.2E+00	AF020528.1	NT	Magnaporthe grisea Class IV chitin synthase (chs4) gene, complete cds
4403	16988	29432	4.5	2.2E+00	D67071.1	NT	Rat gene for regucalcin, exon1 (non-coding exon)
4403	16988	29433	4.5	2.2E+00	D67071.1	NT	Rat gene for regucalcin, exon1 (non-coding exon)
5545	18177	30591	12.27	2.2E+00	O88307	SWISSPROT	SORTILIN-RELATED RECEPTOR PRECURSOR (SORTING PROTEIN-RELATED RECEPTOR CONTAINING LDLR CLASS A REPEATS) (MSORLA) (SORLA-1) (LOW-DENSITY LIPOPROTEIN RECEPTOR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LDLR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LR11) (>
5545	18177	30592	12.27	2.2E+00	O88307	SWISSPROT	SORTILIN-RELATED RECEPTOR PRECURSOR (SORTING PROTEIN-RELATED RECEPTOR CONTAINING LDLR CLASS A REPEATS) (MSORLA) (SORLA-1) (LOW-DENSITY LIPOPROTEIN RECEPTOR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LDLR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LR11) (>
6016	18635	31373	0.95	2.2E+00	BE927220.1	EST_HUMAN	RC3-CT0254-300800-022-e06 CT0254 Homo sapiens cDNA
6016	18635	31374	0.95	2.2E+00	BE927220.1	EST_HUMAN	RC3-CT0254-300800-022-e06 CT0254 Homo sapiens cDNA
6212	18822	31593	9.1	2.2E+00	BE250383.1	EST_HUMAN	600943401T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2959777 3'
6495	19096	31880	4.32	2.2E+00	Q00335	SWISSPROT	MINOR VIRION STRUCTURAL PROTEIN MU-2
6709	19303	32107	3.04	2.2E+00	P51459	SWISSPROT	INSULIN-LIKE GROWTH FACTOR II PRECURSOR (IGF-II) (SOMATOMEDIN A)
7037	18057		3.58	2.2E+00	AA594574.1	EST_HUMAN	ni95b02.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1058379 3'
7358	19884	32747	0.9	2.2E+00	AA137027.1	EST_HUMAN	zn97f04.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:566143 5'
7602	20115	32992	25.23	2.2E+00	AA449012.1	EST_HUMAN	zx05g10.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:785634 5'
8046	20588	33494	0.65	2.2E+00	BE301560.1	EST_HUMAN	bb17h12.x1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2963207 3' similar to gb:D45836 Mouse mRNA for nuclear pore-targeting-complex component of (MOUSE);
8046	20588	33495	0.65	2.2E+00	BE301560.1	EST_HUMAN	bb17h12.x1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2963207 3' similar to gb:D45836 Mouse mRNA for nuclear pore-targeting-complex component of (MOUSE);
9265	21791		12.17	2.2E+00	BE741678.1	EST_HUMAN	601594733F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948561 5'
9488	24793		2.57	2.2E+00	Q04706	SWISSPROT	TRANSPOSON TY1 PROTEIN A
9956	22461	35443	1.96	2.2E+00	AI290373.1	EST_HUMAN	qm69b03.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1893965 3' similar to gb:Y00433 GLUTATHIONE PEROXIDASE (HUMAN);

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9966	22461	35444	1.96	2.2E+00	AI290373.1	EST_HUMAN	qm69b03.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1893965 3' similar to gb:Y00433 GLUTATHIONE PEROXIDASE (HUMAN);
10008	22503	35494	3.7	2.2E+00	BF246782.1	EST_HUMAN	601855591F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075391 5'
10353	22847	35841	2.99	2.2E+00	AF183416.1	NT	Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog mRNA, complete cds
11316	23014	36023	4.01	2.2E+00	P07911	SWISSPROT	UROMODULIN PRECURSOR (TAMM-HORSFALL URINARY GLYCOPROTEIN) (THP)
11482	23932	37003	4.23	2.2E+00	P10407	SWISSPROT	EARLY E1A 28 KD PROTEIN
595	15419	25699	6.28	2.1E+00	AF132612.2	NT	Mus musculus pre-T cell receptor alpha gene, enhancer region and upstream region
3648	16251		0.65	2.1E+00	AW449366.1	EST_HUMAN	UI-H-BI3-aki-e-08-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2734550 3'
6281	18869		0.85	2.1E+00	P75357	SWISSPROT	HYPOTHETICAL PROTEIN MG302 HOMOLOG
6899	19633	32471	3.38	2.1E+00	O70159	SWISSPROT	ALPHA-2-HS-GLYCOPROTEIN PRECURSOR (FETUIN-A)
7110	19450	32266	5.13	2.1E+00	N29575.1	EST_HUMAN	yy08a10.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:270618 3' similar to gb:M55654 TRANSCRIPTION INITIATION FACTOR TFIID (HUMAN);
8434	20974		2.27	2.1E+00	AU123630.1	EST_HUMAN	AU123630 NT2RM2 Homo sapiens cDNA clone NT2RM2000671 5'
10454	22948		0.58	2.1E+00	Y10284.1	NT	H.sapiens TRAF1 gene, putative promoter region
1238	13836	26352	1.3	2.0E+00	AF180527.1	NT	Homo sapiens p22Dokdel (DOKDEL) mRNA, complete cds
1238	13836	26353	1.3	2.0E+00	AF180527.1	NT	Homo sapiens p22Dokdel (DOKDEL) mRNA, complete cds
1380	13973	26501	0.92	2.0E+00	AF204927.1	NT	Oryctolegus cuniculus Na+,K+-ATPase beta 1 subunit mRNA, complete cds
1619	14212		2.89	2.0E+00	P25582	SWISSPROT	PUTATIVE RRNA METHYLTRANSFERASE SPB1
2194	14770	27343	3.69	2.0E+00	Z78279.1	NT	R.norvegicus mRNA for collagen alpha1 type I
2194	14770	27344	3.69	2.0E+00	Z78279.1	NT	R.norvegicus mRNA for collagen alpha1 type I
4176	16767	29215	1.9	2.0E+00	AW664496.1	EST_HUMAN	hi13c05.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2972168 3' similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, LIVER (HUMAN);
4176	16767	29216	1.9	2.0E+00	AW664496.1	EST_HUMAN	hi13c05.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2972168 3' similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, LIVER (HUMAN);
7552	20071		0.77	2.0E+00	P07566	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: NUCLEOCAPSID PROTEIN C; MEMBRANE GLYCOPROTEINS E1 AND E2]
7967	20509	33415	3.56	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds
7967	20509	33416	3.56	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds
7967	20509	33417	3.56	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds
8853	21392	34314	3.62	2.0E+00	F31500.1	EST_HUMAN	HSPD22703 HM3 Homo sapiens cDNA clone s4000117B08
12295	24946	30822	7.77	2.0E+00	5834843	NT	Gallus gallus mitochondrion, complete genome
5784	18409	31124	6.89	1.9E+00	6754389	NT	Mus musculus inositol 1,4,5-triphosphate receptor 1 (Itpr1), mRNA
5784	18409	31125	6.89	1.9E+00	6754389	NT	Mus musculus inositol 1,4,5-triphosphate receptor 1 (Itpr1), mRNA
6249	18858	31630	1.2	1.9E+00	BE969695.1	EST_HUMAN	601679636F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949881 5'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6760	19353		1.02	1.9E+00	AW845689.1	EST_HUMAN	MR0-CT0063-071099-002-g02 CT0063 Homo sapiens cDNA
6845	19435		2.31	1.9E+00	Q63627	SWISSPROT	CTD-BINDING SR-LIKE PROTEIN RA4
8396	20936	33858	2.16	1.9E+00	P02467	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
8396	20936	33859	2.16	1.9E+00	P02467	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
8593	21132		2.45	1.9E+00	BF360206.1	EST_HUMAN	CM3-MT0114-010900-323-h12 MT0114 Homo sapiens cDNA
8825	21364		1.35	1.9E+00	O51781	SWISSPROT	ARGININE DEIMINASE (ADI) (ARGININE DIHYDROLASE) (AD)
9548	22048	35009	0.6	1.9E+00	AA669125.1	EST_HUMAN	ab94a04.s1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:854574 3' similar to contains Alu repetitive element;contains element L1 L1 repetitive element ;
10456	22950	35959	0.52	1.9E+00	AF248269.1	NT	Homo sapiens gag-pro-pol precursor protein gene, partial cds
3128	15742	28211	1.88	1.8E+00	P21004	SWISSPROT	PROTEIN B8 PRECURSOR
3154	15768	28234	2.42	1.8E+00	U04356.1	NT	Synechococcus sp. PCC7942 copper transporting P-ATPase (ctaA) and ATP synthase epsilon subunit (atpE) genes, complete cds
3154	15768	28235	2.42	1.8E+00	U04356.1	NT	Synechococcus sp. PCC7942 copper transporting P-ATPase (ctaA) and ATP synthase epsilon subunit (atpE) genes, complete cds
6027	18646		2.02	1.8E+00	P18502	SWISSPROT	HEDGEHOG RECEPTOR (PATCHED PROTEIN)
6253	18862	31634	2.02	1.8E+00	BF311999.1	EST_HUMAN	601897854F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127364 5'
6532	19132		1.53	1.8E+00	BF683327.1	EST_HUMAN	602139470F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4298272 5'
6838	19428	32244	1.35	1.8E+00	BF305652.1	EST_HUMAN	601893489F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4139038 5'
7119	19459	32274	1.08	1.8E+00	P21249	SWISSPROT	MAJOR ANTIGEN
8060	20602	33512	0.81	1.8E+00	P11369	SWISSPROT	RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]
8060	20602	33513	0.81	1.8E+00	P11369	SWISSPROT	RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]
8788	21327	34252	2.12	1.8E+00	O43281	SWISSPROT	EMBRYONAL FYN-ASSOCIATED SUBSTRATE (HEFS)
9102	21638	34577	0.63	1.8E+00	R31042.1	EST_HUMAN	yh72c08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:135278 5'
9186	21703	34645	0.8	1.8E+00	AW880004.1	EST_HUMAN	QV0-OT0030-070300-148-a03 OT0030 Homo sapiens cDNA
9763	22261	35244	0.87	1.8E+00	P27050	SWISSPROT	CHITINASE D PRECURSOR
10183	22678		3.78	1.8E+00	AF111849.1	NT	Homo sapiens PRO0530 mRNA, complete cds
10447	22941		0.85	1.8E+00	P44325	SWISSPROT	CYTIDINE DEAMINASE (CYTIDINE AMINOHYDROLASE) (CDA)
12075	24915		6.85	1.8E+00	AF314254.1	NT	Chlamydomonas reinhardtii alternative oxidase 1 (AOX1) gene, nuclear gene encoding mitochondrial protein
12183	24403		4.96	1.8E+00	9506404	NT	Rattus norvegicus Actin-related protein complex 1b (Arpc1b), mRNA
12478	24815	30790	1.38	1.8E+00	BF212412.1	EST_HUMAN	601813714F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4048251 5'

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1147	13750	26259	2.08	1.7E+00	Q60114	SWISSPROT	LEVANSUCRASE (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE)
2311	14883	27458	2.37	1.7E+00	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
2411	14979	27554	1.29	1.7E+00	AI141067.1	EST_HUMAN	oz43h05.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1678137 3'
4558	17141	29589	0.74	1.7E+00	Q60114	SWISSPROT	LEVANSUCRASE (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE)
5797	18422	31137	1.65	1.7E+00	BE063546.1	EST_HUMAN	CM0-BT0282-171299-127-e05 BT0282 Homo sapiens cDNA
5797	18422	31138	1.65	1.7E+00	BE063546.1	EST_HUMAN	CM0-BT0282-171299-127-e05 BT0282 Homo sapiens cDNA
6168	18780	31545	3.35	1.7E+00	Q9TTR8	SWISSPROT	COUP TRANSCRIPTION FACTOR 1 (COUP-TF1) (COUP-TF I)
7270	19798	32654	1.33	1.7E+00	Q03703	SWISSPROT	HYPOTHETICAL 38.0 KD PROTEIN IN CAT2-AMD1 INTERGENIC REGION
7270	19798	32655	1.33	1.7E+00	Q03703	SWISSPROT	HYPOTHETICAL 38.0 KD PROTEIN IN CAT2-AMD1 INTERGENIC REGION
7306	19834	32693	1.83	1.7E+00	P20393	SWISSPROT	ORPHAN NUCLEAR RECEPTOR NR1D1 (V-ERBA RELATED PROTEIN EAR-1) (REV-ERBA-ALPHA)
7796	20339	33247	0.96	1.7E+00	AF021335.1	NT	Mus musculus T cell receptor gamma locus, TCR gamma 2 and gamma 4 gene clusters
7976	20518	33425	1.34	1.7E+00	6755715	NT	Mus musculus T-cell acute lymphocytic leukemia 1 (Tal1), mRNA
8006	20548	33452	0.57	1.7E+00	BF530630.1	EST_HUMAN	602071917F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4214669 5'
8479	21018	33933	0.61	1.7E+00	AF245513.1	NT	Hippoglossus hippoglossus interferon inducible Mx protein (Mx) mRNA, complete cds
8562	21101		2.08	1.7E+00	BF308000.1	EST_HUMAN	601894255F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4140084 5'
8638	21177	34096	0.49	1.7E+00	X69063.1	NT	M.musculus Ank-1 mRNA for erythroid ankyrin
8638	21177	34097	0.49	1.7E+00	X69063.1	NT	M.musculus Ank-1 mRNA for erythroid ankyrin
9076	24792	34545	2.25	1.7E+00	O60479	SWISSPROT	HOMEOBOX PROTEIN DLX-3
9076	24792	34546	2.25	1.7E+00	O60479	SWISSPROT	HOMEOBOX PROTEIN DLX-3
9524	22024		1.65	1.7E+00	AF161380.1	NT	Homo sapiens HSPC262 mRNA, partial cds
11467	23917	36985	2.16	1.7E+00	W22424.1	EST_HUMAN	67B7 Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional
12030	24320	30993	1.52	1.7E+00	AI678443.1	EST_HUMAN	tu82d07.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2257549 3' similar to contains MSR1.t1 MSR1 repetitive element ;
12558	24659	30873	1.79	1.7E+00	AI198573.1	EST_HUMAN	qf50b01.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1753417 3' similar to contains L1.t1 L1 repetitive element ;
2078	14658	27229	21.82	1.6E+00	AF199339.1	NT	Homo sapiens lens epithelium-derived growth factor gene, alternatively spliced, complete cds
2087	14668	27238	4.3	1.6E+00	AF077374.1	NT	Homo sapiens small proline-rich protein (SPRR3) gene, exons 1, 2, and 3 and complete cds
2093	14673	27243	1.04	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
2323	14894		1.13	1.6E+00	X98373.1	NT	B.napus gene encoding endo-polygalacturonase
2988	15604	28084	1.5	1.6E+00	W58426.1	EST_HUMAN	zd25f01.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:341689 5' similar to gb:D29805 N-ACETYLLACTOSAMINE SYNTHASE (HUMAN);
4104	16698		7.23	1.6E+00	BF570077.1	EST_HUMAN	602186095T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4310591 3'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4444	17030	29470	1.11	1.6E+00	AF155827.1	NT	Homo sapiens proliferation-associated SNF2-like protein (SMARCA6) mRNA, complete cds
4444	17030	29471	1.11	1.6E+00	AF155827.1	NT	Homo sapiens proliferation-associated SNF2-like protein (SMARCA6) mRNA, complete cds
5145	17715	30145	0.6	1.6E+00	AF075394.1	NT	Uroteuthis chinensis cytochrome c oxidase subunit I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds
5145	17715	30146	0.6	1.6E+00	AF075394.1	NT	Uroteuthis chinensis cytochrome c oxidase subunit I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds
5243	17807	30228	2.2	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
5243	17807	30229	2.2	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
5993	18613	31347	1.95	1.6E+00	L04808.1	NT	Brachydanio rerio MHC class II DA-beta-2*01 gene, 3' end
6072	18689	31434	0.92	1.6E+00	AF005831.1	NT	Homo sapiens transglutaminase type I (Tgasel) gene, promoter region
6596	19193	31998	0.93	1.6E+00	BF380703.1	EST_HUMAN	IL2-UT0073-060900-145-E02 UT0073 Homo sapiens cDNA
6811	19402	32218	1.07	1.6E+00	AW294881.1	EST_HUMAN	UI-H-BI2-ehr-b-04-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone IMAGE:2727511 3'
7293	19821	32680	2.32	1.6E+00	BE697267.1	EST_HUMAN	RC0-CT0415-200700-032-c10 CT0415 Homo sapiens cDNA
7973	20515		1.09	1.6E+00	Q46378	SWISSPROT	VIRULENCE FACTOR MVIN HOMOLOG
8320	20861	33786	3.24	1.6E+00	AJ297131.1	NT	Mus musculus SIL, MAP_17, CYP_a, SCL & CYP_b genes
8831	21370	34294	0.95	1.6E+00	11437222	NT	Homo sapiens hypothetical protein PRO0971 (PRO0971), mRNA
8831	21370	34295	0.95	1.6E+00	11437222	NT	Homo sapiens hypothetical protein PRO0971 (PRO0971), mRNA
9381	24790	33221	3.16	1.6E+00	X52046.1	NT	M.musculus COL3A1 gene for collagen alpha-I
9381	24790	33222	3.16	1.6E+00	X52046.1	NT	M.musculus COL3A1 gene for collagen alpha-I
9649	22148	35119	1.34	1.6E+00	T41290.1	EST_HUMAN	ph6b6_19/1TV Outward Alu-primed hncDNA library Homo sapiens cDNA clone ph6b6_19/1TV
10052	22547	35541	0.52	1.6E+00	AF121361.1	NT	Drosophila melanogaster signal transducting adaptor protein (STAM), serine threonine kinase lal (IAL), and zinc finger protein (DNZ1) genes, complete cds
10088	22583	35575	0.92	1.6E+00	AW835644.1	EST_HUMAN	QV4-LT0016-090200-100-d07 LT0016 Homo sapiens cDNA
10088	22583	35576	0.92	1.6E+00	AW835644.1	EST_HUMAN	QV4-LT0016-090200-100-d07 LT0016 Homo sapiens cDNA
10246	22741	35731	0.49	1.6E+00	AF037352.1	NT	Mus musculus T cell receptor gamma locus, TCR gamma 1 and gamma 3 gene clusters
10650	23182	36196	1.59	1.6E+00	P54817	SWISSPROT	CAPSID PROTEIN P40 [CONTAINS: ASSEMBLIN (PROTEASE) ; CAPSID ASSEMBLY PROTEIN]
10686	23216	36228	1.56	1.6E+00	P54817	SWISSPROT	CAPSID PROTEIN P40 [CONTAINS: ASSEMBLIN (PROTEASE) ; CAPSID ASSEMBLY PROTEIN]
10723	18689	31434	6.41	1.6E+00	AF005831.1	NT	Homo sapiens transglutaminase type I (Tgasel) gene, promoter region
11552	24000	37072	2.92	1.6E+00	AF104313.1	NT	Homo sapiens unknown mRNA
35	12714	25173	4.02	1.5E+00	U53449.1	NT	Rattus norvegicus jun dimerization protein 2 (jdp-2) mRNA, complete cds
252	12912	25396	2.17	1.5E+00	AE002201.2	NT	Chlamydophila pneumoniae AR39, section 32 of 94 of the complete genome
649	13272		1.98	1.5E+00	6752961	NT	Mus musculus a disintegrin and metalloproteinase domain (ADAM) 15 (metargidin) (Adam15), mRNA

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1960	14544	27101	2.55	1.5E+00	AF275265.1	NT	Mus musculus receptor protein tyrosine phosphatase-rho (Ptprt) gene, exons 10 and 11 and partial cds
2454	15021	27592	2.13	1.5E+00	AJ131402.1	NT	Potato virus A RNA complete genome, isolate U
2556	15120	27690	1.83	1.5E+00	6678350	NT	Mus musculus T-cell lymphoma invasion and metastasis 1 (Tiam1), mRNA
3172	15021	27592	1.54	1.5E+00	AJ131402.1	NT	Potato virus A RNA complete genome, isolate U
3421	16029	28510	0.7	1.5E+00	AE001945.1	NT	Deinococcus radiodurans R1 section 82 of 229 of the complete chromosome 1
5903	18525	31250	0.94	1.5E+00	AI655301.1	EST_HUMAN	tt12f10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240587 3' similar to TR:O00237 O00237 HKF-1. ;
5903	18525	31251	0.94	1.5E+00	AI655301.1	EST_HUMAN	tt12f10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240587 3' similar to TR:O00237 O00237 HKF-1. ;
6538	19137	31930	2.68	1.5E+00	R17879.1	EST_HUMAN	yg10e02.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:31693 5'
7182	19714		1.37	1.5E+00	BE785356.1	EST_HUMAN	601478745F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3881555 5'
7214	19745	32599	20.84	1.5E+00	P47179	SWISSPROT	HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR
7214	19745	32600	20.84	1.5E+00	P47179	SWISSPROT	HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR
7383	19909	32774	1.02	1.5E+00	AA889259.1	EST_HUMAN	ak26f10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1407115 3'
8065	20607	33519	0.85	1.5E+00	BE887446.1	EST_HUMAN	601509586F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3911181 5'
8578	21117	34037	1.1	1.5E+00	K02138.1	NT	Mouse germline IgM chain gene, mu-delta region
8946	21484		0.53	1.5E+00	AB038516.1	NT	Homo sapiens hGPIb alpha gene for platelet glycoprotein Ib alpha, complete cds
9061	21598	34528	0.54	1.5E+00	BF217818.1	EST_HUMAN	601862662F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4095135 5'
9404	21913	34862	0.9	1.5E+00	R81928.1	EST_HUMAN	yj03h01.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:147697 5'
9553	22053	35016	1.12	1.5E+00	AW375697.1	EST_HUMAN	QV3-CT0192-261099-008-d09 CT0192 Homo sapiens cDNA
9774	22272	35257	5.97	1.5E+00	BF376754.1	EST_HUMAN	RC0-TN0078-150900-034-g05 TN0078 Homo sapiens cDNA
9965	22460		1.47	1.5E+00	BF337944.1	EST_HUMAN	602035771F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183865 5'
10098	22593	35585	2.95	1.5E+00	AA017689.1	EST_HUMAN	ze38g06.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:361306 5'
10098	22593	35586	2.95	1.5E+00	AA017689.1	EST_HUMAN	ze38g06.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:361306 5'
11277	23730	36785	4.1	1.5E+00	AL134197.1	EST_HUMAN	DKFZp547P243_s1 547 (synonym: hfbr1) Homo sapiens cDNA clone DKFZp547P243 3'
11416	23867		9.57	1.5E+00	X07380.1	NT	Maize mitochondrial tRNA-Ser gene and tRNA-Phe pseudogene
12022	25010	30615	1.59	1.5E+00	D63480.1	NT	Human mRNA for KIAA0146 gene, partial cds
12255	24465		4.99	1.5E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
32	12711	25169	1.8	1.4E+00	7661685	NT	Homo sapiens DKFZP586M0122 protein (DKFZP586M0122), mRNA
32	12711	25170	1.8	1.4E+00	7661685	NT	Homo sapiens DKFZP586M0122 protein (DKFZP586M0122), mRNA
1774	14364	26909	1.32	1.4E+00	H19859.1	EST_HUMAN	yn57e03.r1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:172540 5'
2316	14888		0.98	1.4E+00	AF053357.1	NT	Helicobacter pylori glutamine synthetase (glnA) gene, complete cds
2372	14942		7.8	1.4E+00	U67922.1	NT	Ovis aries prion protein gene, complete cds

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2693	15250	27820	1.45	1.4E+00	X74463.1	NT	Human papillomavirus type 7 genomic DNA
2802	15354	27922	2.79	1.4E+00	AF064564.2	NT	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds
2802	15354	27923	2.79	1.4E+00	AF064564.2	NT	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds
3376	15985		0.68	1.4E+00	5453733	NT	Homo sapiens Mad4 homolog (MAD4) mRNA
4342	16929	29369	1.14	1.4E+00	AW900455.1	EST_HUMAN	CM0-NN1005-140300-286-h06 NN1005 Homo sapiens cDNA
4342	16929	29370	1.14	1.4E+00	AW900455.1	EST_HUMAN	CM0-NN1005-140300-286-h06 NN1005 Homo sapiens cDNA
4685	17267		1.78	1.4E+00	BF581547.1	EST_HUMAN	602156687F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4297556 5'
5575	18206	30657	1.76	1.4E+00	AW054976.1	EST_HUMAN	wt45g07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2510460 3'
5719	18345		5.04	1.4E+00	AB032983.1	NT	Homo sapiens mRNA for KIAA1157 protein, partial cds
6425	19026	31811	2.73	1.4E+00	Q13472	SWISSPROT	DNA TOPOISOMERASE III ALPHA
6437	25116		4.4	1.4E+00	AB020712.1	NT	Homo sapiens mRNA for KIAA0905 protein, complete cds
6544	19143	31936	2.32	1.4E+00	Q92777	SWISSPROT	SYNAPSIN II
6544	19143	31937	2.32	1.4E+00	Q92777	SWISSPROT	SYNAPSIN II
6583	19181	31981	0.67	1.4E+00	11096333	NT	Mus musculus WW domain binding protein 11 (Wbp11-pending), mRNA
6911	19570	32398	0.77	1.4E+00	AW893057.1	EST_HUMAN	CM3-NN0006-300300-132-b12 NN0006 Homo sapiens cDNA
7330	19857	32720	2.31	1.4E+00	AJ133269.1	NT	Homo sapiens caveolin-1/-2 locus, Contig1, D7S522, genes CAV2 (exons 1, 2a, and 2b), CAV1 (exons 1 and 2)
7343	19870	32735	1.1	1.4E+00	AW467760.1	EST_HUMAN	he23f05.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2919873 3' similar to contains Alu repetitive element;
8277	20818		0.68	1.4E+00	P07683	SWISSPROT	GLUCOAMYLASE PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)
8729	21268		4.01	1.4E+00	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
9023	21560	34487	2.13	1.4E+00	R20459.1	EST_HUMAN	yg33f12.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:34345 5'
9125	21660	34603	3.72	1.4E+00	BE064667.1	EST_HUMAN	RC1-BT0313-301299-012-f05 BT0313 Homo sapiens cDNA
9158	21693	34637	0.58	1.4E+00	AF134844.1	NT	Sceloporus undulatus ornithine transcarbamylase (OTC) mRNA, complete cds
10109	22604	35594	0.77	1.4E+00	BF575545.1	EST_HUMAN	602133135F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288137 5'
10150	22645	35637	0.67	1.4E+00	BE145374.1	EST_HUMAN	IL5-HT0198-291099-008-C04 HT0198 Homo sapiens cDNA
10150	22645	35638	0.67	1.4E+00	BE145374.1	EST_HUMAN	IL5-HT0198-291099-008-C04 HT0198 Homo sapiens cDNA
10418	22912	35912	1.11	1.4E+00	D63441.1	NT	Pandorina colemaniae chloroplast rbcL gene for ribulose bisphosphate carboxylase, partial cds
10418	22912	35913	1.11	1.4E+00	D63441.1	NT	Pandorina colemaniae chloroplast rbcL gene for ribulose bisphosphate carboxylase, partial cds
10948	23463	36485	2.16	1.4E+00	AA195528.1	EST_HUMAN	zr36e09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:665512 5' similar to contains element MER22 repetitive element ;

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11104	23614	36654	6.28	1.4E+00	AB006682.1	NT	Homo sapiens APECED mRNA for AIRE-1, complete cds
11283	23736	36791	3.92	1.4E+00	BE962107.2	EST_HUMAN	601655184R1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845805 3'
11283	23736	36792	3.92	1.4E+00	BE962107.2	EST_HUMAN	601655184R1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845805 3'
11304	23797	36855	3.19	1.4E+00	U30790.1	NT	Pneumocystis carinii f. sp. ratti guanine nucleotide binding protein alpha subunit (pcg1) gene, complete cds
11304	23797	36856	3.19	1.4E+00	U30790.1	NT	Pneumocystis carinii f. sp. ratti guanine nucleotide binding protein alpha subunit (pcg1) gene, complete cds
11865	24935		2.43	1.4E+00	AL161500.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 12
12267	25108		2.38	1.4E+00	11545836	NT	Homo sapiens cutaneous T-cell lymphoma tumor antigen se70-2 (SE70-2), mRNA
596	13225		1.38	1.3E+00	Z73640.1	NT	M.mucedo gene encoding 4-Dihydromethyl-trisporate dehydrogenase
935	13548	26085	2.33	1.3E+00	AJ271192.1	NT	Cantharellus sp. partial 25S rRNA gene, isolate Tibet
1168	13770		22.19	1.3E+00	Y19213.1	NT	Homo sapiens putative psiHbA pseudogene for hair keratin, exons 2 to 7
1340	13935	26456	13.67	1.3E+00	4507998	NT	Homo sapiens zinc finger protein 157 (HZF22) (ZNF157) mRNA
1340	13935	26457	13.67	1.3E+00	4507998	NT	Homo sapiens zinc finger protein 157 (HZF22) (ZNF157) mRNA
1400	13994		1.05	1.3E+00	U61730.2	NT	Coix lacryma-jobi dihydrodipicolinate synthase (dapA) gene, complete cds
1653	14245		2.35	1.3E+00	AE002338.2	NT	Chlamydia muridarum, section 66 of 85 of the complete genome
2285	14859		1.1	1.3E+00	AB030447.1	NT	Cyprinus carpio MRPb and MASPb genes for mannose-binding lectin-associated serine protease (MASP) and MASP-related protein, complete cds
2586	15149		0.97	1.3E+00	BE966735.2	EST_HUMAN	601661233R1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915945 3'
2966	15581	28060	0.66	1.3E+00	6755621	NT	Mus musculus alpha-spectrin 1, erythroid (Spna1), mRNA
3657	16260	28732	0.91	1.3E+00	AF016494.1	NT	Fugu rubripes gamma-aminobutyric acid receptor beta subunit gene, partial cds; 55kd erythrocyte membrane protein (P55), synaptic vesicle-associated integral membrane protein (VAMP-1), procollagen C-proteinase enhancer protein (PCOLCE) genes, complete c>
4713	15581	28060	1.31	1.3E+00	6755621	NT	Mus musculus alpha-spectrin 1, erythroid (Spna1), mRNA
5184	17749	30178	0.92	1.3E+00	AJ252087.1	NT	Plasmodium reichenowi partial 83/AMA-1 gene for apical membrane antigen 1
5184	17749	30179	0.92	1.3E+00	AJ252087.1	NT	Plasmodium reichenowi partial 83/AMA-1 gene for apical membrane antigen 1
5705	18331	30835	1.06	1.3E+00	P19732	SWISSPROT	PHENOL HYDROXYLASE P3 PROTEIN (PHENOL 2-MONOOXYGENASE P3 COMPONENT)
6169	18781	31546	7.47	1.3E+00	AW362834.1	EST_HUMAN	PM0-CT0289-291199-004-f08 CT0289 Homo sapiens cDNA
6169	18781	31547	7.47	1.3E+00	AW362834.1	EST_HUMAN	PM0-CT0289-291199-004-f08 CT0289 Homo sapiens cDNA
6549	19147	31943	1.24	1.3E+00	M33496.1	NT	D.melanogaster no-on-transient A gene product, complete cds
6847	19437		0.71	1.3E+00	Q00156	SWISSPROT	HYPOTHETICAL GENE 64 PROTEIN
6926	19585	32415	0.85	1.3E+00	M13918.2	NT	Homo sapiens fibronectin receptor alpha-subunit precursor (ITGA5) mRNA, partial cds
7033	19567	32394	1.01	1.3E+00	BE538819.1	EST_HUMAN	601061420F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3447965 5'

WO 01/57277

PCT/US01/00669

210/526

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7157	19689	32633	0.79	1.3E+00	BE243571.1	EST_HUMAN	TCBAP1D0959 Pediatric pre-B cell acute lymphoblastic leukemia Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP0959
7481	20003	32868	3.97	1.3E+00	P24540	SWISSPROT	ACYLPHOSPHATASE, ORGAN-COMMON TYPE ISOZYMES A AND B (ACYLPHOSPHATE PHOSPHOHYDROLASE)
8239	20780	33701	2.06	1.3E+00	AJ009912.1	NT	Sus scrofa plp gene
8384	20924	33844	2.54	1.3E+00	BE963379.2	EST_HUMAN	601657145R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866195 3'
8496	21035	33956	0.89	1.3E+00	BE974280.1	EST_HUMAN	601680250R2 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:3950532 3'
8644	21183		1.57	1.3E+00	9910247	NT	Homo sapiens GL004 protein (GL004), mRNA
8725	21264	34184	0.88	1.3E+00	AI927629.1	EST_HUMAN	wo85a07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2462100 3'
9073	21610	34540	0.48	1.3E+00	H42881.1	EST_HUMAN	yo68c03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:183076 3'
9073	21610	34541	0.48	1.3E+00	H42881.1	EST_HUMAN	yo68c03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:183076 3'
9434	21960		4.54	1.3E+00	AF042084.1	NT	Homo sapiens heparan glucosaminyl N-deacetylase/N-sulfotransferase-2 gene, complete cds
9443	21969	34917	2.12	1.3E+00	X72019.1	NT	S.alba phr-1 mRNA for photolyase
9443	21969	34918	2.12	1.3E+00	X72019.1	NT	S.alba phr-1 mRNA for photolyase
9542	22042	35003	1.1	1.3E+00	AF059250.1	NT	Homo sapiens lipoxygenase (ALOX12B) mRNA, complete cds
9588	22088	35052	1.62	1.3E+00	O00754	SWISSPROT	LYSOSOMAL ALPHA-MANNOSIDASE PRECURSOR (MANNOSIDASE, ALPHA B) (LYSOSOMAL ACID ALPHA-MANNOSIDASE) (LAMAN)
9665	22164	35137	1.21	1.3E+00	AI927629.1	EST_HUMAN	wo85a07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2462100 3'
9740	22238	35218	0.83	1.3E+00	AJ223962.1	NT	Lactococcus lactis cremoris NCDO-inv1 chromosomal inversion junction DNA
9740	22238	35219	0.83	1.3E+00	AJ223962.1	NT	Lactococcus lactis cremoris NCDO-inv1 chromosomal inversion junction DNA
9780	22278	35263	3.85	1.3E+00	BE963379.2	EST_HUMAN	601657145R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866195 3'
10114	22609	35600	1.25	1.3E+00	AE004392.1	NT	Vibrio cholerae chromosome II, section 49 of 93 of the complete chromosome
10130	22625	35615	2.41	1.3E+00	M29953.1	NT	Campylobacter jejuni kanamycin phosphotransferase (aphA-7) gene, complete cds
10476	22970		0.65	1.3E+00	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
10504	22998		0.52	1.3E+00	8923637	NT	Homo sapiens hypothetical protein FLJ20707 (FLJ20707), mRNA
10507	23001	36008	0.48	1.3E+00	H42881.1	EST_HUMAN	yo68c03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:183076 3'
10507	23001	36009	0.48	1.3E+00	H42881.1	EST_HUMAN	yo68c03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone IMAGE:183076 3'
10573	23108		4.66	1.3E+00	Q14117	SWISSPROT	DIHYDROPYRIMIDINASE (DHPASE) (HYDANTOINASE) (DHP)
10785	23309	36316	2.3	1.3E+00	P25299	SWISSPROT	MRNA 3'-END PROCESSING PROTEIN RNA15
10807	23330	36342	2.17	1.3E+00	Z18892.2	NT	Mus musculus desmin gene
11215	23718		1.87	1.3E+00	AW274791.1	EST_HUMAN	xp09e03.x1 NCI_CGAP_HN9 Homo sapiens cDNA clone IMAGE:2739868 3'
11414	23865	36926	3.09	1.3E+00	D42042.1	NT	Human mRNA for KIAA0085 gene, partial cds
11488	23937	37007	3.09	1.3E+00	Z98682.1	NT	Bacillus subtilis genomic DNA 23.9kB fragment
12011	24312		3.63	1.3E+00	AF187873.1	NT	Cavia porcellus inwardly-rectifying potassium channel Kir2.2 (KCNJ12) gene, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12192	24423	30949	3.47	1.3E+00	BF348043.1	EST_HUMAN	602023185F1 NCI_CGAP_Bm67 Homo sapiens cDNA clone IMAGE:4158452 5'
12204	24826		1.76	1.3E+00	P33464	SWISSPROT	E1 GLYCOPROTEIN PRECURSOR (MATRIX GLYCOPROTEIN) (MEMBRANE GLYCOPROTEIN)
12303	24500		2.08	1.3E+00	AF187035.1	NT	Sturnira lilium cytochrome b gene, complete cds; mitochondrial gene for mitochondrial product
12673	24904		1.25	1.3E+00	AL163283.2	NT	Homo sapiens chromosome 21 segment HS21C083
678	13302	25784	9.75	1.2E+00	AA676246.1	EST_HUMAN	zl22d08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431535 3'
856	13472	25983	1.04	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
856	13472	25984	1.04	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
856	13472	25985	1.04	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
911	13524		1.9	1.2E+00	8924234	NT	Homo sapiens hypothetical protein PRO3077 (PRO3077), mRNA
1203	13803	26316	4.87	1.2E+00	AF080245.2	NT	Elaeis oleifera sesquiterpene synthase mRNA, complete cds
1247	13844	26361	1.3	1.2E+00	AJ252242.1	NT	pea seed-borne mosaic virus complete genome
1247	13844	26362	1.3	1.2E+00	AJ252242.1	NT	pea seed-borne mosaic virus complete genome
2054	14635	27206	53.59	1.2E+00	AF140631.1	NT	Homo sapiens G-protein coupled receptor 14 (GPR14) gene, complete cds
2417	14985	27559	1.53	1.2E+00	AF156495.1	NT	Homo sapiens post-synaptic density 95 (DLG4) gene, complete cds
3144	15758	28224	1.16	1.2E+00	AB020681.1	NT	Homo sapiens mRNA for KIAA0874 protein, partial cds
3201	15813	28287	7.17	1.2E+00	AL161563.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 63
3201	15813	28288	7.17	1.2E+00	AL161563.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 63
3325	15935		3.43	1.2E+00	P54910	SWISSPROT	CONJUGAL TRANSFER PROTEIN TRBE PRECURSOR
3399	16007	28489	0.57	1.2E+00	AF188740.1	NT	Homo sapiens LHX3 gene, intron 2
3774	16374	28839	8.66	1.2E+00	U75902.1	NT	Mus musculus subtilisin-like serine protease LPC (PC7) gene, exons 1 to 9, partial cds
4058	16655	29121	1.87	1.2E+00	BF373570.1	EST_HUMAN	MR0-FT0175-050900-203-g06_1 FT0175 Homo sapiens cDNA
4386	16007	28489	1.12	1.2E+00	AF188740.1	NT	Homo sapiens LHX3 gene, intron 2
4570	17153		2.09	1.2E+00	M87060.1	NT	Rattus rattus cardiac AE3 gene, exons 1-23
4621	17204	29653	1.08	1.2E+00	AL161509.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 21
4659	17241	29695	1.5	1.2E+00	AF156495.1	NT	Homo sapiens post-synaptic density 95 (DLG4) gene, complete cds
4690	17272		9.41	1.2E+00	Y09200.1	NT	T.pinnatum chloroplast rbcL gene, partial
4791	16008		0.77	1.2E+00	M81779.1	NT	G.gallus T-cadherin mRNA, complete cds
5629	18258	30729	1.06	1.2E+00	U20760.1	NT	Human extracellular calcium-sensing receptor mRNA, complete cds
5743	18369	31077	2.27	1.2E+00	AW813276.1	EST_HUMAN	MR3-ST0191-140200-013-c05 ST0191 Homo sapiens cDNA
6034	18653	31395	0.72	1.2E+00	AF016052.1	NT	Homo sapiens zinc finger protein ZNF191 (ZNF191) gene, complete cds
6300	18908	31679	2.17	1.2E+00	X74885.1	NT	D.hydei ayt repeat cluster DNA, fragment D
6361	18965	31743	3.98	1.2E+00	BE003113.1	EST_HUMAN	QV4-BN0090-270400-190-a03 BN0090 Homo sapiens cDNA
6433	19036	31822	1.43	1.2E+00	X89084.1	NT	C.glutamicum pta gene and ackA gene
6433	19036	31823	1.43	1.2E+00	X89084.1	NT	C.glutamicum pta gene and ackA gene

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6475	19076	31859	34.96	1.2E+00	AA759254.1	EST_HUMAN	ah84g12.s1 Soares_testis_NHT Homo sapiens cDNA clone 1322374 3'
6629	19225	32030	2.25	1.2E+00	AW813276.1	EST_HUMAN	MR3-ST0191-140200-013-c05 ST0191 Homo sapiens cDNA
6995	19493	32314	1.18	1.2E+00	AB029010.1	NT	Homo sapiens mRNA for KIAA1087 protein, partial cds
7007	19505	32324	2.8	1.2E+00	AJ002141.1	NT	Mus musculus DSPP gene
7300	19828		0.8	1.2E+00	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
7417	24782	32808	1.59	1.2E+00	AV734585.1	EST_HUMAN	AV734585 cdA Homo sapiens cDNA clone cdAAFH03 5'
7646	20158	33045	2.84	1.2E+00	X74207.1	NT	L.lactis pyrD and pyrF genes
8504	21043	33964	3.05	1.2E+00	AB033030.1	NT	Homo sapiens mRNA for KIAA1204 protein, partial cds
8597	21136	34051	0.69	1.2E+00	P38427	SWISSPROT	ALPHA,ALPHA-TREHALOSE-PHOSPHATE SYNTHASE [UDP-FORMING] 123 KD SUBUNIT (TREHALOSE-6-PHOSPHATE SYNTHASE) (UDP-GLUCOSE-GLUCOSEPHOSPHATE GLUCOSYLTRANSFERASE)
8809	21348		0.53	1.2E+00	7706271	NT	Homo sapiens CGI-30 protein (LOC51611), mRNA
8955	21493	34416	2.03	1.2E+00	AW377210.1	EST_HUMAN	MR2-CT0222-201099-001-e07 CT0222 Homo sapiens cDNA
9319	21833	34783	2.92	1.2E+00	Z32850.1	NT	R.communis gene for pyrophosphate-dependent phosphofructokinase beta subunit
9523	22023	34981	1.86	1.2E+00	D11745.1	EST_HUMAN	HUMHM01A01 Liver HepG2 cell line. Homo sapiens cDNA clone hm01a01
9844	22342	35324	3.47	1.2E+00	X56832.1	NT	H.sapiens ENO3 gene for muscle specific enolase
10229	22724		0.67	1.2E+00	AB009666.1	NT	Homo sapiens klotho gene, exon 1
11224	23755	36813	2.19	1.2E+00	AW817817.1	EST_HUMAN	PM0-ST0264-161199-001-d01 ST0264 Homo sapiens cDNA
11262	23790		6.64	1.2E+00	BE160761.1	EST_HUMAN	PM1-HT0422-160200-007-g10 HT0422 Homo sapiens cDNA
11331	23029	36038	3.76	1.2E+00	U50147.1	NT	Rattus norvegicus synapse-associated protein 102 mRNA, complete cds
11976	24907	30712	32.4	1.2E+00	AL163203.2	NT	Homo sapiens chromosome 21 segment HS21C003
11998	24304		2.11	1.2E+00	AP001515.1	NT	Bacillus halodurans genomic DNA, section 9/14
489	13122	25608	1.19	1.1E+00	D86980.1	NT	Human mRNA for KIAA0227 gene, partial cds
1799	14389	26934	1.48	1.1E+00	AW995393.1	EST_HUMAN	QV0-BN0042-170300-163-g12 BN0042 Homo sapiens cDNA
2617	15179	27746	1.09	1.1E+00	AF067124.1	NT	Wheat yellow mosaic virus RNA1 270 kDa precursor protein gene, complete cds
3373	15981	28458	9.32	1.1E+00	AL163213.2	NT	Homo sapiens chromosome 21 segment HS21C013
3373	15981	28459	9.32	1.1E+00	AL163213.2	NT	Homo sapiens chromosome 21 segment HS21C013
3533	16138	28620	0.84	1.1E+00	8922641	NT	Homo sapiens hypothetical protein FLJ10749 (FLJ10749), mRNA
3639	16242	28718	1.06	1.1E+00	AI808360.1	EST_HUMAN	wf54h11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2359481 3' similar to SW:P531_HUMAN Q12888 P53-BINDING PROTEIN 53BP1 ;
3781	16381	28845	1.41	1.1E+00	AE003886.1	NT	Xylella fastidiosa, section 32 of 229 of the complete genome
3781	16381	28846	1.41	1.1E+00	AE003886.1	NT	Xylella fastidiosa, section 32 of 229 of the complete genome
3889	16488		0.61	1.1E+00	X85374.1	NT	H.parahaemolyticus hphIM(A), hphIM(C), hphIR and menB genes
4016	16614	29087	0.67	1.1E+00	8922641	NT	Homo sapiens hypothetical protein FLJ10749 (FLJ10749), mRNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4093	16688	29145	0.65	1.1E+00	6755205	NT	Mus musculus proteasome (prosome, macropain) subunit, beta type 7 (Psmb7), mRNA
4295	16881		7.81	1.1E+00	5835331	NT	R.unicornis complete mitochondrial genome
5131	17703	30137	3.54	1.1E+00	U18466.1	NT	African swine fever virus, complete genome
5132	17704	30138	0.96	1.1E+00	AJ271740.1	NT	Drosophila melanogaster D-Titin gene, exons 1-37
5201	17766	30190	1	1.1E+00	U34740.1	NT	Emericella nidulans sterigmatocystin biosynthetic gene cluster: (stcA), (stcB), (stcC), (stcE), (aflR), (stcF), (stcI), (stcJ), (stcK), (stcL), (stcO), (stcQ), (stcS), (stcT), (stcU), (stcV) and (stcW) genes, complete cds
5230	17794	30213	0.96	1.1E+00	X78425.1	NT	E.faecalis pbp5 gene
5409	17966		1.04	1.1E+00	AE003869.1	NT	Xylella fastidiosa, section 15 of 229 of the complete genome
5510	18143	30555	1.52	1.1E+00	6978530	NT	Rattus norvegicus Aquaporin 4 (Aqp4), mRNA
5798	18423	31139	19.98	1.1E+00	BE960184.1	EST_HUMAN	601652776R1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3825835 3'
5815	18439	31161	3.23	1.1E+00	AI138582.1	EST_HUMAN	qd85c03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1736260 3'
6242	18851	31621	1.25	1.1E+00	11418739	NT	Homo sapiens solute carrier family 6 (neurotransmitter transporter), member 14 (SLC6A14), mRNA
6420	19023	31807	0.71	1.1E+00	AF197861.1	NT	Macgregoria pulchra cytochrome b gene, complete cds; mitochondrial gene for mitochondrial product
6539	19138	31931	0.71	1.1E+00	R06037.1	EST_HUMAN	ye89e03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:124924 5'
6817	19408	32225	0.7	1.1E+00	AJ404004.1	NT	Mus musculus mRNA for ER protein 58 (ER58 gene)
7338	19865	32729	0.78	1.1E+00	X55981.1	NT	Maize mRNA for enolase (2-phospho-D-glycerate hydrolase)
7501	20023	32886	2.08	1.1E+00	Z72338.1	NT	Herpes simplex virus type 1 (strain KOS) UL41 gene
7501	20023	32887	2.08	1.1E+00	Z72338.1	NT	Herpes simplex virus type 1 (strain KOS) UL41 gene
7521	20041	32910	9.72	1.1E+00	AL161588.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 84
7580	24787	32973	0.99	1.1E+00	11967960	NT	Mus musculus silent mating type information regulation 2, (S.cerevisiae, homolog)-like (Sir2l), mRNA
8074	20616	33530	2.8	1.1E+00	BF693998.1	EST_HUMAN	602082582F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4246828 5'
8163	20704	33620	0.75	1.1E+00	AI478339.1	EST_HUMAN	tm39h11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2160549 3'
8672	21211	34130	0.75	1.1E+00	AB003088.1	NT	Acetabularia caliculus mitochondrial COXI-like gene
8749	21288	34208	0.78	1.1E+00	S80750.1	NT	VH=anti-cytomegalovirus glycoprotein B antibody 4D4 heavy chain variable region [human, mRNA Partial, 375 nt]
9358	20297		0.68	1.1E+00	BE384876.1	EST_HUMAN	601276278F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617418 5'
9546	22046	35007	0.68	1.1E+00	AJ245772.1	NT	Mus musculus mRNA for stretch responsive muscle (X-chromosome) protein (Srmx gene)
9599	22099		0.81	1.1E+00	Y12227.1	NT	Arabidopsis thaliana DNA, 24 kb surrounding PFL locus
9687	22186	35160	0.84	1.1E+00	L76301.1	NT	Yersinia pseudotuberculosis psaE, psaF, adhesin (psaA), chaperone (psaB), and usher (psaC) genes, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9747	22245	35226	1.59	1.1E+00	AB023151.1	NT	Homo sapiens mRNA for KIAA0934 protein, partial cds
9850	22348	35330	4.82	1.1E+00	AL161515.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 27
9909	22406	35381	19.39	1.1E+00	6754021	NT	Mus musculus guanine nucleotide binding protein (G protein), gamma 3 subunit (Gng3), mRNA
10395	22889	35883	1	1.1E+00	P73769	SWISSPROT	DNA MISMATCH REPAIR PROTEIN MUTS
10530	23067	36079	2.93	1.1E+00	11067364	NT	Homo sapiens KIAA0626 gene product (KIAA0626), mRNA
10586	23121		4.06	1.1E+00	AF068942.1	NT	Klebsormidium fluitans cytochrome c oxidase subunit 2 (cox2) gene, mitochondrial gene encoding mitochondrial protein, partial cds
10978	18026		5.28	1.1E+00	8922973	NT	Homo sapiens hypothetical protein FLJ11280 (FLJ11280), mRNA
10983	23497	36526	3.76	1.1E+00	AF012862.1	NT	Petroselinum crispum cytosolic glucose-6-phosphate dehydrogenase 1 (cG6PDH1) mRNA, complete cds
10983	23497	36527	3.76	1.1E+00	AF012862.1	NT	Petroselinum crispum cytosolic glucose-6-phosphate dehydrogenase 1 (cG6PDH1) mRNA, complete cds
11234	23765	36822	6.02	1.1E+00	AI809699.1	EST_HUMAN	wf76e11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2361548 3'
11946	24275		1.82	1.1E+00	P07866	SWISSPROT	LOW TEMPERATURE ESSENTIAL PROTEIN
12051	24335	30997	2.25	1.1E+00	AF216696.1	NT	Taenia solium Immunogenic protein Ts76 mRNA, partial cds
12184	24903		1.64	1.1E+00	AF234169.1	NT	Dictyostelium discoideum isopentenyl pyrophosphate isomerase (Dipi) mRNA, complete cds
103	12779		3.22	1.0E+00	U23808.1	NT	Xenopus laevis rhodopsin gene, complete cds
118	12789	25271	3.48	1.0E+00	D88425.1	NT	Cavia cobaya mRNA for serine/threoine kinase, complete cds
443	13076		2.14	1.0E+00	AB021684.1	NT	Marchantia polymorpha genes for 26S rRNA, 5S rRNA, 18S rRNA, 5.8S rRNA and 26S rRNA
602	13231	25704	1.53	1.0E+00	AJ251660.1	NT	Girardia tigrina mRNA for homeodomain transcription factor (so gene)
705	13326	25813	7.12	1.0E+00	AL163218.2	NT	Homo sapiens chromosome 21 segment HS21C018
707	13328		0.89	1.0E+00	AF125984.1	NT	Aedes aegypti mucin-like protein MUC1 mRNA, complete cds
1429	15441		1.73	1.0E+00	X80416.1	NT	V.carteri Algal-CAM mRNA
1794	14384	26929	0.91	1.0E+00	AB006531.1	NT	Plautia stali intestine virus RNA for nonstructural polyprotein, capsid protein precursor, complete cds
2526	15090	27662	1.2	1.0E+00	P48355	SWISSPROT	DNA GYRASE SUBUNIT B
2526	15090	27663	1.2	1.0E+00	P48355	SWISSPROT	DNA GYRASE SUBUNIT B
2900	15517	27986	4.47	1.0E+00	P24008	SWISSPROT	3-OXO-5-ALPHA-STEROID 4-DEHYDROGENASE 1 (STEROID 5-ALPHA-REDUCTASE 1) (SR TYPE 1)
2900	15517	27987	4.47	1.0E+00	P24008	SWISSPROT	3-OXO-5-ALPHA-STEROID 4-DEHYDROGENASE 1 (STEROID 5-ALPHA-REDUCTASE 1) (SR TYPE 1)
2994	15610		0.83	1.0E+00	O14226	SWISSPROT	HYPOTHETICAL 67.9 KD PROTEIN C6F12.08C IN CHROMOSOME I
3232	15844	28324	0.91	1.0E+00	AA628453.1	EST_HUMAN	af26g08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:1032830 3' similar to WP:C42D8.3 CE04204 ;contains element MER22 MER22 repetitive element ;

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3659	12779		0.78	1.0E+00	U23808.1	NT	Xenopus laevis rhodopsin gene, complete cds
3744	16345	28813	1.55	1.0E+00	AJ223816.1	NT	Agaricus bisporus mRNA for tyrosinase
4144	16738	29189	1.5	1.0E+00	AF223391.1	NT	Homo sapiens calcium channel alpha1E subunit (CACNA1E) gene, exons 7-49, and partial cds, alternatively spliced
4362	16949		0.64	1.0E+00	8922245	NT	Homo sapiens hypothetical protein FLJ10139 (FLJ10139), mRNA
5165	17734		17.2	1.0E+00	D10852.1	NT	Rattus norvegicus mRNA for N-acetylglucosaminyltransferase III, complete cds
5339	17900		0.6	1.0E+00	AF200817.1	NT	Pilot whale morbillivirus phosphoprotein (P) gene, partial cds
5432	18010	30394	1	1.0E+00	AB039022.1	NT	Oncorhynchus mykiss stl1 mRNA for rhamnose binding lectin STL1, complete cds
5486	18120	30527	2.56	1.0E+00	Z97022.1	NT	Hordeum vulgare gene encoding cysteine proteinase
6013	18633	31368	4.54	1.0E+00	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
6013	18633	31369	4.54	1.0E+00	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
6111	18727	31480	1.22	1.0E+00	Z97341.2	NT	Arabidopsis thaliana DNA chromosome 4, ESSA 1 FCA contig fragment No. 6
6263	18871	31641	4.41	1.0E+00	P04501	SWISSPROT	FIBER PROTEIN
6269	18877	31645	1.56	1.0E+00	AW452782.1	EST_HUMAN	UI-H-BI3-alx-d-09-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:3068969 3'
6615	19212	32018	1.79	1.0E+00	U75902.1	NT	Mus musculus subtilisin-like serine protease LPC (PC7) gene, exons 1 to 9, partial cds
6662	19258	32062	0.83	1.0E+00	AF104869.1	NT	Homo sapiens cell cycle protein (PA2G4) gene, exons 2 though 5
6742	19336		1.5	1.0E+00	P46506	SWISSPROT	SRB-11 PROTEIN
6874	19608	32442	1.27	1.0E+00	Y11204.1	NT	V.carteri gene encoding volvoxopsin
7192	19724	32573	1.22	1.0E+00	S52770.1	NT	insulin-like growth factor-binding protein 4 [cattle, pulmonary artery endothelial cells, mRNA, 2028 nt]
7493	20016		8.58	1.0E+00	P20273	SWISSPROT	B-CELL RECEPTOR CD22 PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL ADHESION MOLECULE) (BL-CAM)
7697	20206	33093	1.36	1.0E+00	AF192531.1	NT	Homo sapiens endothelin-converting enzyme 2 (ECE2) mRNA, complete cds
7710	20219	33107	7.92	1.0E+00	AA775191.1	EST_HUMAN	ac79b08.s1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:868791 3'
7902	20444	33349	1.49	1.0E+00	BE868267.1	EST_HUMAN	601443950F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848005 5'
7902	20444	33350	1.49	1.0E+00	BE868267.1	EST_HUMAN	601443950F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848005 5'
8084	17734		1.28	1.0E+00	D10852.1	NT	Rattus norvegicus mRNA for N-acetylglucosaminyltransferase III, complete cds
8291	20832	33753	2.1	1.0E+00	Q02207	SWISSPROT	PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8291	20832	33754	2.1	1.0E+00	Q02207	SWISSPROT	PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]
8413	20953		0.85	1.0E+00	P51784	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 11) (DEUBIQUITINATING ENZYME 11)
8447	20987	33902	0.48	1.0E+00	Q9Y5T5	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16) (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M)
8447	20987	33903	0.48	1.0E+00	Q9Y5T5	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16) (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M)
8475	24791		2.17	1.0E+00	BE147331.1	EST_HUMAN	RC1-HT0229-181099-011-e06 HT0229 Homo sapiens cDNA
8513	21052	33974	1.06	1.0E+00	U42720.2	NT	Simian immunodeficiency virus Gag protein (gag) gene, complete cds; Pol protein (pol) gene, partial cds; and Vif protein (vif), Vpr protein (vpr), Tat protein (tat), Rev protein (rev), Vpu protein (vpu), Env protein (env), and Nef protein (nef) genes, >
8659	21198	34116	1.07	1.0E+00	M38427.1	NT	Human immunodeficiency virus type 1 (HIV-1), isolate SF33,
9195	21712	34655	2.05	1.0E+00	BE907592.1	EST_HUMAN	601497581F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3899421 5'
9402	21911	34860	1.34	1.0E+00	6753429	NT	Mus musculus chloride channel calcium activated 1 (Clca1), mRNA
9402	21911	34861	1.34	1.0E+00	6753429	NT	Mus musculus chloride channel calcium activated 1 (Clca1), mRNA
9528	22028	34987	2.06	1.0E+00	AV689554.1	EST_HUMAN	AV689554 GKC Homo sapiens cDNA clone GKCCYA11 5'
9534	22034	34993	1.33	1.0E+00	U44952.1	NT	Xenopus laevis zona pellucida C glycoprotein precursor (xlZPC) mRNA, complete cds
9534	22034	34994	1.33	1.0E+00	U44952.1	NT	Xenopus laevis zona pellucida C glycoprotein precursor (xlZPC) mRNA, complete cds
9767	22265	35248	0.5	1.0E+00	X15498.1	NT	Human Coronavirus gene for membrane protein
9767	22265	35249	0.5	1.0E+00	X15498.1	NT	Human Coronavirus gene for membrane protein
10021	22516	35510	0.62	1.0E+00	5174562	NT	Homo sapiens MHC binding factor, beta (MHCBFB) mRNA
10021	22516	35511	0.62	1.0E+00	5174562	NT	Homo sapiens MHC binding factor, beta (MHCBFB) mRNA
10105	22600	35592	0.75	1.0E+00	AI077920.1	EST_HUMAN	oy15d07.s1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1665901 3'
10230	22725	35716	4.17	1.0E+00	AV758825.1	EST_HUMAN	AV758825 BM Homo sapiens cDNA clone BMFAWC04 5'
10372	22866	35859	19.78	1.0E+00	AA004982.1	EST_HUMAN	zh94a02.r1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:428906 5'
10372	22866	35860	19.78	1.0E+00	AA004982.1	EST_HUMAN	zh94a02.r1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:428906 5'
10404	22898	35893	0.93	1.0E+00	L11910.1	NT	Human retinoblastoma susceptibility gene exons 1-27, complete cds
10853	23374	36393	1.87	1.0E+00	S90825.1	NT	PBR1=proline-rich protein (intron 3) [human, Genomic, 898 nt]
11587	18120	30527	1.57	1.0E+00	Z97022.1	NT	Hordeum vulgare gene encoding cysteine proteinase
11837	24201		4.85	1.0E+00	P15306	SWISSPROT	THROMBOMODULIN PRECURSOR (FETOMODULIN) (TM)

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12171	24410		3.08	1.0E+00	AW976184.1	EST_HUMAN	EST388293 MAGE resequences, MAGN Homo sapiens cDNA
1616	14209	26742	0.97	9.9E-01	AF245455.1	NT	Drosophila melanogaster regulator of G-protein signalling LOCO III mRNA, complete cds
1616	14209	26743	0.97	9.9E-01	AF245455.1	NT	Drosophila melanogaster regulator of G-protein signalling LOCO III mRNA, complete cds
2664	15222	27794	1.17	9.9E-01	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
3665	16267		0.94	9.9E-01	AF174585.1	NT	Apple mosaic virus RNA 2 putative polymerase gene, complete cds
5816	18440	31162	14.59	9.9E-01	P49657	SWISSPROT	SERINE/THREONINE PROTEIN KINASE MINIBRAIN
6029	18648	31389	0.83	9.9E-01	Q09632	SWISSPROT	PROBABLE OXIDOREDUCTASE ZK1290.5 IN CHROMOSOME II
9185	21702		1.39	9.9E-01	U65667.1	NT	Lycopersicon esculentum putative Mi1 copy 1 nematode-resistance gene
9474	21873		2.61	9.9E-01	Q28642	SWISSPROT	B2 BRADYKININ RECEPTOR (BK-2 RECEPTOR)
10593	23128	36142	1.68	9.9E-01	AJ005029.1	NT	Danio rerio mRNA for Eph-like receptor tyrosine kinase rtk8
549	13160	25658	1.77	9.8E-01	P22567	SWISSPROT	AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS) (NAGS)
2336	14907		0.89	9.8E-01	AJ003108.1	NT	Callithrix jacchus UBE1 gene derived retroposon on the Y chromosome
2827	15379		2.05	9.8E-01	AF174644.1	NT	Xenopus laevis rac GTPase mRNA, complete cds
3869	16467	28930	0.95	9.8E-01	O67551	SWISSPROT	PROBABLE ENDONUCLEASE IV (ENDODEOXYRIBONUCLEASE IV)
3872	16470	28933	0.61	9.8E-01	BE957439.2	EST_HUMAN	601653583R2 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838461 3'
3872	16470	28934	0.61	9.8E-01	BE957439.2	EST_HUMAN	601653583R2 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838461 3'
7250	19779	32634	4.86	9.8E-01	AJ302158.1	NT	Enterobacteriaceae sp. JM983 partial groES gene for GroES-like protein and partial groEL gene for GroEL-like protein, isolate JM983
7250	19779	32635	4.86	9.8E-01	AJ302158.1	NT	Enterobacteriaceae sp. JM983 partial groES gene for GroES-like protein and partial groEL gene for GroEL-like protein, isolate JM983
7641	20153	33038	1.13	9.8E-01	BF034016.1	EST_HUMAN	601456337F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3860049 5'
7641	20153	33039	1.13	9.8E-01	BF034016.1	EST_HUMAN	601456337F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3860049 5'
8853	21192	34110	0.77	9.8E-01	P38652	SWISSPROT	PHOSPHOGLUCOMUTASE (GLUCOSE PHOSPHOMUTASE) (PGM)
10334	22828		0.56	9.8E-01	AA825565.1	EST_HUMAN	od55d04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371847 3'
10874	23395	36410	4.86	9.8E-01	BE258705.1	EST_HUMAN	601110258F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3350750 5'
10874	23395	36411	4.86	9.8E-01	BE258705.1	EST_HUMAN	601110258F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3350750 5'
11597	24040	37109	1.78	9.8E-01	AI680876.1	EST_HUMAN	tx42c10.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272242 3'
12058	24341		1.39	9.8E-01	U52111.2	NT	Homo sapiens X28 region near ALD locus containing dual specificity phosphatase 9 (DUSP9), ribosomal protein L18a (RPL18a), Ca2+/Calmodulin-dependent protein kinase I (CAMKI), creatine transporter (CRTR), CDM protein (CDM), adrenoleukodystrophy protein >
7212	19743	32597	2.51	9.7E-01	U26716.1	NT	Drosophila melanogaster sodium channel protein (para) gene, exons 9,10,11,12 and optional segments b, c, d and e, partial cds
8440	20980	33895	1.7	9.7E-01	AF149112.1	NT	Triticum aestivum stripe rust resistance protein Yr10 (Yr10) gene, complete cds
8446	20986	33901	1.28	9.7E-01	M90544.1	NT	Salmonella typhimurium adenine-methyltransferase (mod) and restriction endonuclease (res)

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11049	23562		5.23	9.7E-01	BF511209.1	EST_HUMAN	UI-H-BI4-aoi-e-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone IMAGE:3085140 3'
12658	24729		2.92	9.7E-01	AL114281.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
4531	17115	29559	0.58	9.6E-01	AF197925.1	NT	Bromus inermis putative cytosolic phosphoglucomutase (pgm1) mRNA, complete cds
4531	17115	29560	0.58	9.6E-01	AF197925.1	NT	Bromus inermis putative cytosolic phosphoglucomutase (pgm1) mRNA, complete cds
4557	17140	29588	1.71	9.6E-01	AW799674.1	EST_HUMAN	PM2-UM0053-240300-005-f12 UM0053 Homo sapiens cDNA
5928	18550	31276	3.9	9.6E-01	Z70556.1	NT	Parvovirus B19 DNA, patient C, genome position 2448-2994
5928	18550	31277	3.9	9.6E-01	Z70556.1	NT	Parvovirus B19 DNA, patient C, genome position 2448-2994
8331	20872		1.23	9.6E-01	X95275.1	NT	P.falciparum complete gene map of plastid-like DNA (IR-A)
8785	21324	34248	0.47	9.6E-01	L81138.1	NT	Rattus norvegicus (strain R21) Rps2r gene, complete cds
9020	21557	34485	0.62	9.6E-01	AF229843.1	NT	Mus musculus WNT-2 gene, partial cds; putative ankyrin-related protein and cystic fibrosis transmembrane conductance regulator (CFTR) genes, section 1 of 2 of the complete cds; and unknown gene
10966	23481	36507	1.81	9.6E-01	AF041427.1	NT	Homo sapiens ribosomal protein s4 Y isoform gene, complete cds
11395	23847	36912	5.18	9.6E-01	AV752605.1	EST_HUMAN	AV752605 NPD Homo sapiens cDNA clone NPDBAG06 5'
11395	23847	36913	5.18	9.6E-01	AV752605.1	EST_HUMAN	AV752605 NPD Homo sapiens cDNA clone NPDBAG06 5'
11733	24138		2.36	9.6E-01	11421722	NT	Homo sapiens centrosomal protein 2 (CEP2), mRNA
12388	24983	30609	2.8	9.6E-01	U91423.1	NT	Sphyrna tiburo NADH dehydrogenase subunit 2 (NADH2) gene, mitochondrial gene encoding mitochondrial protein, partial cds
2515	15079	27651	1.02	9.5E-01	7705591	NT	Homo sapiens CGI-125 protein (LOC51003), mRNA
2691	15248	27817	1.2	9.5E-01	Q02934	SWISSPROT	ENDOGLUCANASE I PRECURSOR (EGI) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE I)
3850	16448	28909	1.89	9.5E-01	BE902340.1	EST_HUMAN	601675639F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958473 5'
3850	16448	28910	1.89	9.5E-01	BE902340.1	EST_HUMAN	601675639F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958473 5'
8931	21469	34387	0.63	9.5E-01	AI190162.1	EST_HUMAN	qd57d07.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1733581 3'
9034	21571	34500	1.07	9.5E-01	AW861102.1	EST_HUMAN	RC1-CT0295-241199-011-b02 CT0295 Homo sapiens cDNA
11123	23631	36674	1.71	9.5E-01	BF218771.1	EST_HUMAN	601885163F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4103630 5'
11326	23024	36033	1.59	9.5E-01	AW293799.1	EST_HUMAN	UI-H-BI2-ahp-f-03-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone IMAGE:2727677 3'
3235	15847		1.8	9.4E-01	AF165990.1	NT	Bartonella clarridgeiae RNA polymerase beta subunit (rpoB) gene, partial cds
3254	15866		2.47	9.4E-01	AF080595.1	NT	Pimpinella brachycarpa zinc finger protein (ZFP1) mRNA, complete cds
8799	21338	34265	0.88	9.4E-01	M90724.1	NT	Human Fc-gamma-receptorIIA (FCGR2A) gene, exon 4
1768	14358		0.95	9.3E-01	AF242382.1	NT	Homo sapiens phytanoyl-CoA hydroxylase (PHYH) gene, exon 5
2662	15220	27792	1.09	9.3E-01	BE071172.1	EST_HUMAN	RC5-BT0503-271199-011-B01 BT0503 Homo sapiens cDNA
4107	16701	29154	0.92	9.3E-01	M20219.1	NT	Bovine papillomavirus type 2, complete genome
4107	16701	29155	0.92	9.3E-01	M20219.1	NT	Bovine papillomavirus type 2, complete genome

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5778	18403	31119	1.41	9.3E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
5858	18481	31204	3.69	9.3E-01	L36189.1	NT	Spodoptera frugiperda methylenetetrahydrofolate dehydrogenase mRNA, complete cds
8011	20553	33456	1.62	9.3E-01	AA847040.1	EST_HUMAN	oe09b03.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1385357
8748	21287		1.13	9.3E-01	AF061981.1	NT	Xenopus laevis CCCH zinc finger protein C3H-2 (C3H-2) mRNA, complete cds
8867	21406	34330	1.01	9.3E-01	AL161534.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 34
12506	24629	30893	1.87	9.3E-01	11440298	NT	Homo sapiens inositol 1,4,5-triphosphate receptor, type 2 (ITPR2), mRNA
12515	24634		2	9.3E-01	AF271207.1	NT	Aedes triseriatus putative large subunit ribosomal protein rpL34 mRNA, complete cds
3276	15887	28369	3.99	9.2E-01	BE622702.1	EST_HUMAN	601441338T1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3916184 3'
5004	17577		0.62	9.2E-01	BF129973.1	EST_HUMAN	601817814F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4041363 5'
5894	18516		1.41	9.2E-01	7106410	NT	Mus musculus solute carrier family 30 (zinc transporter), member 4 (Slc30a4), mRNA
6140	18754	31512	4.4	9.2E-01	BF037586.1	EST_HUMAN	601461153F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864661 5'
9578	22078	35042	1.31	9.2E-01	AL161565.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 65
9663	22162	35135	1.15	9.2E-01	6671677	NT	Mus musculus carbonic anhydrase 4 (Car4), mRNA
10166	22661	35656	3.47	9.2E-01	11430963	NT	Homo sapiens lysosomal apyrase-like protein 1 (LALP1), mRNA
10314	22808	35800	1.58	9.2E-01	BF593251.1	EST_HUMAN	7o58e06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3578219 3' similar to SW:NU5M_TRYBB P04540 NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 ;
10526	23063	36074	1.75	9.2E-01	BE563811.1	EST_HUMAN	601334943F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3688714 5'
11569	24016	37085	2.27	9.2E-01	BF132402.1	EST_HUMAN	601820312F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4052018 5'
1666	14259	26793	4.89	9.1E-01	T96675.1	EST_HUMAN	ye52f01.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:121369 3' similar to contains Alu repetitive element;
2169	14746		2.38	9.1E-01	8923056	NT	Homo sapiens hypothetical protein FLJ20048 (FLJ20048), mRNA
3239	15851	28331	0.93	9.1E-01	T26418.1	EST_HUMAN	AB200G8R Infant brain, LLNL array of Dr. M. Soares 1NIB Homo sapiens cDNA clone LLAB200G8 5'
3239	15851	28332	0.93	9.1E-01	T26418.1	EST_HUMAN	AB200G8R Infant brain, LLNL array of Dr. M. Soares 1NIB Homo sapiens cDNA clone LLAB200G8 5'
6315	18922	31699	1.42	9.1E-01	L36033.1	NT	Human pre-B cell stimulating factor homologue (SDF1b) mRNA, complete cds
6630	19226	32031	2.82	9.1E-01	Q81704	SWISSPROT	INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H3 PRECURSOR (ITI HEAVY CHAIN H3)
7577	20093	32970	15.95	9.1E-01	AA806623.1	EST_HUMAN	ob71g08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336862 3'
7719	20227	33115	3.12	9.1E-01	U72995.1	NT	Rattus norvegicus Rab3 GDP/GTP exchange protein mRNA, complete cds
12093	24976		33.14	9.1E-01	AF050113.1	NT	Homo sapiens uncoupling protein-3 (UCP3) gene, complete cds
3241	15853	28335	0.81	9.0E-01	7661625	NT	Homo sapiens DKFZP564M2423 protein (DKFZP564M2423), mRNA
3401	16010		0.64	9.0E-01	AL161515.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 27
4468	17054	29498	1.44	9.0E-01	AF099810.1	NT	Homo sapiens neurexin III-alpha gene, partial cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7424	19948	32814	0.78	9.0E-01	L42547.1	NT	Danio rerio LIM class homeodomain protein (lim5) mRNA, complete cds
7450	19974		1.64	9.0E-01	D38621.1	NT	Xenopus laevis gene for aldolase, complete cds
9271	21797	34746	0.54	9.0E-01	AF086761.1	NT	Danio rerio semaphorin Z1a mRNA, complete cds
9744	22242	35223	0.47	9.0E-01	U39702.1	NT	Mycoplasma genitalium section 24 of 51 of the complete genome
5875	18497	31222	2.49	8.9E-01	AF026198.1	NT	Fugu rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and calcium channel alpha-1 subunit>
6396	18999		1.27	8.9E-01	X60986.1	NT	Rabbit MHC fragment RLA-DF DNA
8152	20693		0.47	8.9E-01	AF260225.1	NT	Homo sapiens TESTIN 2 and TESTIN 3 genes, complete cds, alternatively spliced
8365	20905	33824	1.04	8.9E-01	AF259667.1	NT	Oithona nana cytochrome-c oxidase subunit I (coxI) gene, partial cds; mitochondrial gene for mitochondrial product
11616	24058	37122	2.59	8.9E-01	AE003944.1	NT	Xylella fastidiosa, section 90 of 229 of the complete genome
11927	24262		5.33	8.9E-01	AE002186.2	NT	Chlamydophila pneumoniae AR39, section 21 of 94 of the complete genome
4640	17222	29676	2.1	8.8E-01	O26350	SWISSPROT	PUTATIVE F420-DEPENDENT NADP REDUCTASE
5576	18207	30658	0.7	8.8E-01	AF310617.1	NT	Pseudorabies virus Ea glycoprotein M gene, complete cds
10960	23475	36500	3.82	8.8E-01	Z28337.1	NT	M.aeruginosa (HUB 5-2-4) DNA from plasmid PMA1
11749	25087		2.27	8.8E-01	D90911.1	NT	Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643
490	13123	25609	1.48	8.7E-01	AF106953.2	NT	Homo sapiens SOS1 (SOS1) gene, partial cds
2446	15013	27585	1.13	8.7E-01	5901893	NT	Homo sapiens AT-binding transcription factor 1 (ATBF1), mRNA
2898	15515	27984	5.67	8.7E-01	AA595863.1	EST_HUMAN	nn05f11.s1 NCI_CGAP_Pr4.1 Homo sapiens cDNA clone IMAGE:1076877
4845	17423	29875	0.61	8.7E-01	AF156539.1	NT	Homo sapiens xeroderma pigmentosum complementation group C (XPC) gene, intron 9
4845	17423	29876	0.61	8.7E-01	AF156539.1	NT	Homo sapiens xeroderma pigmentosum complementation group C (XPC) gene, intron 9
5151	17721		3.08	8.7E-01	AF121970.1	NT	Pseudomonas aeruginosa topoisomerase (top), putative transcriptional regulatory protein OhbR (ohbR), ortho-halobenzoate 1,2-dioxygenase beta-ISP protein OhbA (ohbA), OhbC (ohbC), ortho-halobenzoate 1,2-dioxygenase alpha-ISP protein OhbB (ohbB), and put>
7983	20525	33431	0.66	8.7E-01	AW897335.1	EST_HUMAN	RC4-NN0057-120500-013-c07 NN0057 Homo sapiens cDNA
8860	21399	34322	0.75	8.7E-01	AI239456.1	EST_HUMAN	qh36e06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1846786 3'
8860	21399	34323	0.75	8.7E-01	AI239456.1	EST_HUMAN	qh36e06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1846786 3'
9653	22152	35122	1.7	8.7E-01	AE004963.1	NT	Pseudomonas aeruginosa PA01, section 524 of 529 of the complete genome
10205	22700	35693	0.56	8.7E-01	BF570169.1	EST_HUMAN	602185541T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4309906 3'
10205	22700	35694	0.56	8.7E-01	BF570169.1	EST_HUMAN	602185541T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4309906 3'
10711	23239	36254	5.79	8.7E-01	BF363970.1	EST_HUMAN	QV0-NN1021-100800-337-c03 NN1021 Homo sapiens cDNA
11582	24028	37097	4.31	8.7E-01	BF107694.1	EST_HUMAN	601823684R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043564 3'
11582	24028	37098	4.31	8.7E-01	BF107694.1	EST_HUMAN	601823684R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043564 3'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12146	24861		4.44	8.7E-01	AV661898.1	EST_HUMAN	AV661898 GLC Homo sapiens cDNA clone GLCGYG07 3'
500	13132		1.55	8.6E-01	X17012.1	NT	Rat IGFII gene for insulin-like growth factor II
891	13505	26024	8.72	8.6E-01	W69089.1	EST_HUMAN	zd44e03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:343516 5'
2310	14882	27457	1.06	8.6E-01	4503210	NT	Homo sapiens cytochrome P450, subfamily XXVIIA (sterold 27-hydroxylase, cerebrotendinous xanthomatosis), polypeptide 1 (CYP27A1b) mRNA
3681	16282	28750	0.78	8.6E-01	AL161565.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 65
3870	16468	28931	1.38	8.6E-01	U49724.1	NT	Drosophila melanogaster merlin (Dmerlin) mRNA, complete cds
6057	18674	31415	9.06	8.6E-01	X60547.1	NT	Chicken lipoprotein lipase gene
6057	18674	31416	9.06	8.6E-01	X60547.1	NT	Chicken lipoprotein lipase gene
6810	19401	32216	1.88	8.6E-01	AF143732.1	NT	Grus canadensis recombination activating protein 1 (RAG-1) gene, partial cds
6810	19401	32217	1.88	8.6E-01	AF143732.1	NT	Grus canadensis recombination activating protein 1 (RAG-1) gene, partial cds
7868	20410		1.33	8.6E-01	AP001518.1	NT	Bacillus halodurans genomic DNA, section 12/14
7986	20528	33434	0.54	8.6E-01	AF077837.1	NT	Drosophila melanogaster collapsin response mediator protein (CRMP) mRNA, complete cds
9603	22103		0.48	8.6E-01	AE000979.1	NT	Archaeoglobus fulgidus section 128 of 172 of the complete genome
12338	24812		1.73	8.6E-01	AL112162.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
6826	19416	32232	1.32	8.5E-01	AF165214.1	NT	Bacteriophage D3, complete genome
7533	20053	32926	2.38	8.5E-01	BE542612.1	EST_HUMAN	601067107F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3453505 5'
7932	20474	33383	0.51	8.5E-01	AL161572.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 68
8357	20897	33817	0.84	8.5E-01	P06601	SWISSPROT	SEGMENTATION PROTEIN PAIRED
8357	20897	33818	0.84	8.5E-01	P06601	SWISSPROT	SEGMENTATION PROTEIN PAIRED
8441	20981	33896	0.51	8.5E-01	AJ243213.1	NT	Homo sapiens partial 5-HT4 receptor gene, exons 2 to 5
10252	22747	35734	1.38	8.5E-01	AB006799.1	NT	Cyanidium caldarium gene for SigC, complete cds
10252	22747	35735	1.38	8.5E-01	AB006799.1	NT	Cyanidium caldarium gene for SigC, complete cds
12077	24978		3.12	8.5E-01	11418543	NT	Homo sapiens human immunodeficiency virus type I enhancer-binding protein 1 (HIVEP1), mRNA
12084	24355		7.92	8.5E-01	9507008	NT	Rattus norvegicus protein tyrosine phosphatase, non-receptor type 5 (Ptpn5), mRNA
4862	17440	29890	0.62	8.4E-01	AF083975.2	NT	Fowl adenovirus 8, complete genome
5685	24747	30807	3.15	8.4E-01	L78726.1	NT	Human fibroblast growth factor receptor 3 (FGFR3) gene, intron 7
5685	24747	30808	3.15	8.4E-01	L78726.1	NT	Human fibroblast growth factor receptor 3 (FGFR3) gene, intron 7
9868	22365		2.68	8.4E-01	AJ248287.1	NT	Pyrococcus abyssi complete genome; segment 5/6
771	13390	25889	2.48	8.3E-01	M93437.1	NT	Thermus thermophilus cytochrome c-552 (cycA) and CycB (cycB) genes, complete cds
3129	15743	28212	3.26	8.3E-01	AL161506.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 18
3883	16481	28943	0.66	8.3E-01	AB010879.1	NT	Nicotiana tabacum mRNA for chloroplast ribosomal protein L10, complete cds
4084	16680	29140	3.24	8.3E-01	Y19177.1	NT	Streptomyces antibioticus polyketide biosynthetic gene cluster
5473	18107	30426	2.15	8.3E-01	AL161540.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9587	22087		3.14	8.3E-01	AI791952.1	EST_HUMAN	nn01f12.y5 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1076495 5' similar to contains THR.t1 THR repetitive element ;
10019	22514	35507	1.11	8.3E-01	AF098070.1	NT	Drosophila melanogaster Lis1 homolog mRNA, complete cds
10119	22614	35604	3.5	8.3E-01	AF108133.1	NT	Mus musculus neuro-d4 gene, exons 3 through 12 and partial cds
10553	23089	36103	2.92	8.3E-01	AE000903.1	NT	Methanobacterium thermoautotrophicum from bases 1270510 to 1283409 (section 109 of 148) of the complete genome
10571	23106		2.52	8.3E-01	7212472	NT	Phytophthora infestans mitochondrion, complete genome
11183	23688	36735	2.45	8.3E-01	AF020503.1	NT	Homo sapiens FRA3B common fragile region, diadenosine triphosphate hydrolase (FHIT) gene, exon 5
2096	14675	27244	3.23	8.2E-01	AB000489.1	NT	Rattus norvegicus mRNA for RPHO-1, complete cds
2137	14715		1.45	8.2E-01	AF145589.1	NT	Mus musculus trophinin (Tnn) gene, complete cds
3969	16567	29036	1.12	8.2E-01	AB014574.1	NT	Homo sapiens mRNA for KIAA0674 protein, partial cds
4209	16798	29246	0.61	8.2E-01	Z72584.1	NT	S.cerevisiae chromosome VII reading frame ORF YGL062w
4209	16798	29247	0.61	8.2E-01	Z72584.1	NT	S.cerevisiae chromosome VII reading frame ORF YGL062w
5270	17832	30258	1.08	8.2E-01	AB000489.1	NT	Rattus norvegicus mRNA for RPHO-1, complete cds
5420	17977	30385	2.11	8.2E-01	AB028957.1	NT	Homo sapiens mRNA for KIAA1034 protein, partial cds
6871	19605	32439	0.8	8.2E-01	AJ010142.1	NT	Amanita muscaria mRNA for SCIII25 protein
6978	19554	32379	3.18	8.2E-01	AW379433.1	EST_HUMAN	CM4-HT0243-081199-037-e01 HT0243 Homo sapiens cDNA
7313	24779	32700	4.21	8.2E-01	Z12126.1	NT	S.cerevisiae MET, LEU4, and POL1 genes encoding MET4 protein, alpha-isopropylmalate (alpha-IPM) synthetase (partial), and DNA polymerase alpha (partial)
9938	22433	35409	0.63	8.2E-01	AB014530.1	NT	Homo sapiens mRNA for KIAA0630 protein, partial cds
9971	22466	35450	1.67	8.2E-01	AF052659.1	NT	Homo sapiens thioredoxin-related protein mRNA, complete cds
10123	22618	35609	0.59	8.2E-01	AF223888.1	NT	Oncorhynchus tshawytscha isolate T-20 somatolactin precursor gene, exon 1
10123	22618	35610	0.59	8.2E-01	AF223888.1	NT	Oncorhynchus tshawytscha isolate T-20 somatolactin precursor gene, exon 1
10286	22781	35772	3.52	8.2E-01	Q9JI70	SWISSPROT	MCKUSICK-KAUFMAN/BARDET-BIEDL SYNDROMES PUTATIVE CHAPERONIN
10286	22781	35773	3.52	8.2E-01	Q9JI70	SWISSPROT	MCKUSICK-KAUFMAN/BARDET-BIEDL SYNDROMES PUTATIVE CHAPERONIN
11498	23947	37017	3.33	8.2E-01	L10127.1	NT	Molluscum contagiosum virus type 1 ORF1 and ORF2 DNA
11576	24022	37091	8.05	8.2E-01	P10383	SWISSPROT	OVARIAN TUMOR LOCUS PROTEIN
11583	24029	37099	5.68	8.2E-01	H87398.1	EST_HUMAN	yw14d02.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:252195 5' similar to gb:M36072 60S RIBOSOMAL PROTEIN L7A (HUMAN);
12102	24364	30970	1.66	8.2E-01	AJ001261.1	NT	Mus musculus mRNA for NIPSNAP2 protein
2787	15340		1.79	8.1E-01	AF191839.1	NT	Mus musculus TANK binding kinase TBK1 (Tbk1) mRNA, complete cds
3504	16109	28585	2.99	8.1E-01	AF055066.1	NT	Homo sapiens MHC class 1 region
3504	16109	28586	2.99	8.1E-01	AF055066.1	NT	Homo sapiens MHC class 1 region

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5046	17619		0.66	8.1E-01	AF202634.1	NT	Drosophila melanogaster Na/K-ATPase beta subunit isoform 4 (JYbeta2) mRNA, complete cds
6457	19058	31843	0.88	8.1E-01	U16790.1	NT	Mus musculus putative collagen alpha-2 (XI) chain (COL11A2) gene, partial cds
6735	19329	32134	2.54	8.1E-01	Q13491	SWISSPROT	NEURONAL MEMBRANE GLYCOPROTEIN M6-B
6735	19329	32135	2.54	8.1E-01	Q13491	SWISSPROT	NEURONAL MEMBRANE GLYCOPROTEIN M6-B
7852	20394	33298	0.84	8.1E-01	AF022713.2	NT	Drosophila melanogaster putative inorganic phosphate cotransporter (Picot) gene, partial cds; putative sodium channel (Nach) and putative amylase-related protein (Amyrel) genes, complete cds; and putative serine-enriched protein (gprs) gene, partial cd>
7852	20394	33299	0.84	8.1E-01	AF022713.2	NT	Drosophila melanogaster putative inorganic phosphate cotransporter (Picot) gene, partial cds; putative sodium channel (Nach) and putative amylase-related protein (Amyrel) genes, complete cds; and putative serine-enriched protein (gprs) gene, partial cd>
8545	21084	34006	0.92	8.1E-01	AP001517.1	NT	Bacillus halodurans genomic DNA, section 11/14
8545	21084	34007	0.92	8.1E-01	AP001517.1	NT	Bacillus halodurans genomic DNA, section 11/14
8705	21244	34167	1.08	8.1E-01	AW242647.1	EST_HUMAN	xn01h03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2692469 3' similar to SW:LYAR_MOUSE Q08288 CELL GROWTH REGULATING NUCLEOLAR PROTEIN. ;contains MER22.b1 PTR5 repetitive element ;
10032	22527	35522	0.54	8.1E-01	P06425	SWISSPROT	PROBABLE E4 PROTEIN
11356	23810	36869	2.97	8.1E-01	BE938558.1	EST_HUMAN	RC0-TN0080-220800-025-d10 TN0080 Homo sapiens cDNA
11356	23810	36870	2.97	8.1E-01	BE938558.1	EST_HUMAN	RC0-TN0080-220800-025-d10 TN0080 Homo sapiens cDNA
11811	24183	31031	3.32	8.1E-01	AE001711.1	NT	Thermotoga maritima section 23 of 136 of the complete genome
188	12849		4.99	8.0E-01	AJ271510.1	NT	Staphylococcus aureus partiel pta gene for phosphate actyltransferase allele 15
310	12965	25453	7.95	8.0E-01	AJ132772.1	NT	Bos taurus futb and rtlf genes
2080	14661		1.47	8.0E-01	BF530962.1	EST_HUMAN	602072473F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215091 5'
3113	15728	28199	1.24	8.0E-01	AF127897.1	NT	Saimiri boliviensis olfactory receptor (SBO27) gene, partial cds
3354	15962	28439	1.13	8.0E-01	AB006193.1	NT	Mus musculus gene for oviductal glycoprotein, complete cds
3765	16366		1.05	8.0E-01	AL162758.2	NT	Neisseria meningitidis serogroup A strain Z2491 complete genome; segment 7/7
4630	17213	29664	5.65	8.0E-01	X83739.2	NT	G.gallus mRNA for nicotinic acetylcholine receptor (nAChR) beta 3 subunit
5117	17689	30127	1.09	8.0E-01	7657352	NT	Mus musculus myosin IXb (Myo9b), mRNA
7931	20473		2.32	8.0E-01	AW901489.1	EST_HUMAN	RC0-NN1012-270300-021-h06 NN1012 Homo sapiens cDNA
8462	21002	33919	1.17	8.0E-01	Y11095.1	NT	Rice stripe virus RNA 3
479	13112	25602	1.37	7.9E-01	D11476.1	NT	Lymantria dispar nuclear polyhedrosis virus gene for DNA polymerase, complete cds
744	13364		1.05	7.9E-01	AE002130.1	NT	Ureaplasma urealyticum section 31 of 59 of the complete genome
1648	14240		28.9	7.9E-01	AB040885.1	NT	Homo sapiens mRNA for KIAA1452 protein, partial cds
1695	14288		1.11	7.9E-01	U32739.1	NT	Haemophilus influenzae Rd section 54 of 163 of the complete genome
2303	14876	27452	6.76	7.9E-01	AB004816.1	NT	Oryctolagus cuniculus mRNA for mitsugumin29, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2304	14877	27453	5.48	7.9E-01	AF130459.1	NT	Danio rerio Trp4-associated protein Tap1A (tap1A) mRNA, complete cds
3567	16171	28653	2.33	7.9E-01	AF228664.1	NT	Gallus gallus SOX8 transcription factor (SOX8) mRNA, complete cds
4389	16975		0.76	7.9E-01	BE263612.1	EST_HUMAN	601192033F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3535785 5'
4717	17298	29743	1.04	7.9E-01	6753745	NT	Mus musculus embigin (Emb), mRNA
4717	17298	29744	1.04	7.9E-01	6753745	NT	Mus musculus embigin (Emb), mRNA
5315	17877		5.8	7.9E-01	M29930.1	NT	Human insulin receptor (allele 2) gene, exons 14, 15, 16 and 17
6485	19086	31868	0.69	7.9E-01	D38145.1	NT	Human mRNA for prostacyclin synthase, complete cds
8053	20595	33502	2.52	7.9E-01	X90996.1	NT	P.sativum GR gene
9466	21991	34948	4.57	7.9E-01	U01912.1	NT	Giardia lamblia variant-specific surface protein G3M-B (vspG3M-B) mRNA, partial cds
9962	22457	35440	4.27	7.9E-01	P19719	SWISSPROT	SMALL HYDROPHOBIC PROTEIN
10003	22498	35487	0.75	7.9E-01	AV700860.1	EST_HUMAN	AV700860 GKC Homo sapiens cDNA clone GKCDRE12 3'
10405	22899	35894	0.71	7.9E-01	AB000831.1	NT	Streptococcus mutans DNA for sigma 42 protein, dTDP-4-keto-L-rhamnose reductase, complete cds
10886	23407		2.28	7.9E-01	7662471	NT	Homo sapiens KIAA1072 protein (KIAA1072), mRNA
11089	23601	36639	2.72	7.9E-01	P19022	SWISSPROT	NEURAL-CADHERIN PRECURSOR (N-CADHERIN)
909	13522		1.4	7.8E-01	Z43785.1	EST_HUMAN	HSC1KH041 normalized infant brain cDNA Homo sapiens cDNA clone c-1kh04
2314	14888	27461	14	7.8E-01	AW959567.1	EST_HUMAN	EST371637 MAGE resequences, MAGF Homo sapiens cDNA
4811	17389	29840	0.81	7.8E-01	U87305.1	NT	Rattus norvegicus transmembrane receptor Unc5H1 mRNA, complete cds
5189	17754		0.81	7.8E-01	AW753353.1	EST_HUMAN	RC3-CT0254-130100-023-c02 CT0254 Homo sapiens cDNA
6219	18829	31603	2.33	7.8E-01	AF115856.1	NT	Sphenodon punctatus alpha enolase mRNA, partial cds
6367	18971	31750	1.05	7.8E-01	P05231	SWISSPROT	INTERLEUKIN-6 PRECURSOR (IL-6) (B-CELL STIMULATORY FACTOR 2) (BSF-2) (INTERFERON BETA-2) (HYBRIDOMA GROWTH FACTOR)
6589	19186	31988	0.75	7.8E-01	AL445066.1	NT	Thermoplasma acidophilum complete genome, segment 4/5
8428	20968	33881	1.04	7.8E-01	BF108927.1	EST_HUMAN	7I54d05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3525176 3'
9160	21695	34639	1.02	7.8E-01	Y10159.1	NT	D.discoideum racGAP gene
9255	21781	34733	0.53	7.8E-01	4826873	NT	Homo sapiens nucleoporin 214kD (CAIN) (NUP214), mRNA
10031	22526		0.78	7.8E-01	Q25452	SWISSPROT	MUSCLE CALCIUM CHANNEL ALPHA-1 SUBUNIT (MDL-ALPHA1)
12071	24957		2.33	7.8E-01	L29260.1	NT	Arabidopsis thaliana 1-amino-1-cyclopropanecarboxylate synthase (ACS5) gene, complete cds
150	12813	25300	4.65	7.7E-01	AF184345.1	NT	Lycopersicon hirsutum ADP-glucose pyrophosphorylase large subunit (AGP-L1) mRNA, complete cds
755	13374		1.44	7.7E-01	AF050157.1	NT	Mus musculus major histocompatibility locus class II region: major histocompatibility protein class II alpha chain (IAalpha) and major histocompatibility protein class II beta chain (IEbeta) genes, complete cds; butyrophilin-like (NG9), butyrophilin-li>
2737	15292	27860	2.33	7.7E-01	O33915	SWISSPROT	CITRATE SYNTHASE

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3400	16009		0.62	7.7E-01	8393408	NT	Homo sapiens UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNAc-T7) (GALNAC-T7), mRNA
3660	16262	28734	4.78	7.7E-01	AF118085.1	NT	Homo sapiens PRO1975 mRNA, complete cds
4486	17071	29521	3.17	7.7E-01	AF199488.1	NT	Coturnix coturnix japonica sub-species japonica beta-actin mRNA, partial cds
4486	17071	29522	3.17	7.7E-01	AF199488.1	NT	Coturnix coturnix japonica sub-species japonica beta-actin mRNA, partial cds
5749	18375	31084	1.33	7.7E-01	P16553	SWISSPROT	RAFFINOSE INVERTASE (INVERTASE)
5749	18375	31085	1.33	7.7E-01	P16553	SWISSPROT	RAFFINOSE INVERTASE (INVERTASE)
6110	18726	31479	0.8	7.7E-01	R08600.1	EST_HUMAN	yf24b02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:127755 3'
9758	22256	35239	0.82	7.7E-01	AB021134.1	NT	Daphnia magna hemoglobin gene cluster (dhb3, dhb1 and dhb2 genes), complete cds
11957	24280		15.01	7.7E-01	11497621	NT	Archaeoglobus fulgidus, complete genome
4790	17370	29822	19.73	7.6E-01	L27316.1	NT	Oryctolagus cuniculus immunoglobulin VDJ region gene
4790	17370	29823	19.73	7.6E-01	L27316.1	NT	Oryctolagus cuniculus immunoglobulin VDJ region gene
6248	18857	31628	4.81	7.6E-01	AF059510.1	NT	Arabidopsis thaliana 3-methylcrotonyl-CoA carboxylase non-biotinylated subunit (MCCB) mRNA, complete cds
6248	18857	31629	4.81	7.6E-01	AF059510.1	NT	Arabidopsis thaliana 3-methylcrotonyl-CoA carboxylase non-biotinylated subunit (MCCB) mRNA, complete cds
6641	19237	32039	0.7	7.6E-01	P37938	SWISSPROT	MATING-TYPE PROTEIN A-ALPHA Z4
6935	18043	30465	0.95	7.6E-01	AI253399.1	EST_HUMAN	aq14b12.x1 Stanley Frontal NS pool 2 Homo sapiens cDNA clone IMAGE:2030879
6935	18043	30488	0.95	7.6E-01	AI253399.1	EST_HUMAN	aq14b12.x1 Stanley Frontal NS pool 2 Homo sapiens cDNA clone IMAGE:2030879
7113	19453	32269	0.98	7.6E-01	U72487.1	NT	Rattus norvegicus calcium-independent alpha-latrotoxin receptor mRNA, complete cds
8009	20551	33455	1.34	7.6E-01	AF146793.2	NT	Mus musculus neuromedin U precursor (Nmu) gene, partial cds; tPhLP (Tphlp) gene, partial cds; CLOCK (Clock) gene, complete cds; PFT27 (Pft27) gene, complete cds; and H5AR (H5ar) gene, complete cds
8068	20610	33522	1.76	7.6E-01	6857752	NT	Mus musculus advillin (Advil-pending), mRNA
8068	20610	33523	1.76	7.6E-01	6857752	NT	Mus musculus advillin (Advil-pending), mRNA
8267	20808	33727	0.55	7.6E-01	Q01098	SWISSPROT	GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C)
8267	20808	33728	0.55	7.6E-01	Q01098	SWISSPROT	GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C)
8897	21435	34359	0.91	7.6E-01	6753577	NT	Mus musculus cytochrome P450, 2b9, phenobarbitol inducible, type a (Cyp2b9), mRNA
9203	21720	34664	3.33	7.6E-01	P30372	SWISSPROT	MUSCARINIC ACETYLCHOLINE RECEPTOR M2
9203	21720	34665	3.33	7.6E-01	P30372	SWISSPROT	MUSCARINIC ACETYLCHOLINE RECEPTOR M2
11236	23767	36824	2.74	7.6E-01	X86347.1	NT	H.asperse mRNA for neurofilament NF70
11236	23767	36825	2.74	7.6E-01	X86347.1	NT	H.asperse mRNA for neurofilament NF70

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11556	24004		5.74	7.6E-01	AL161592.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 88
11711	24121		6.31	7.6E-01	AB020702.1	NT	Homo sapiens mRNA for KIAA0895 protein, partial cds
539	13170		1.32	7.5E-01	AL163301.2	NT	Homo sapiens chromosome 21 segment HS21C101
610	13238	25712	1.13	7.5E-01	AF020503.1	NT	Homo sapiens FRA3B common fragile region, diadenosine triphosphate hydrolase (FHIT) gene, exon 5
7530	20050	32923	0.74	7.5E-01	AF052730.1	NT	Drosophila melanogaster tyrosine kinase receptor protein (eph) mRNA, complete cds
12027	24318		5.26	7.5E-01	AF163151.2	NT	Homo sapiens dentin sialophosphoprotein precursor (DSPP) gene, complete cds
12522	24638	30897	1.91	7.5E-01	D90907.1	NT	Synechocystis sp. PCC6803 complete genome, 9/27, 1056467-1188885
1169	13771	26279	1.36	7.4E-01	AI598146.1	EST_HUMAN	tn14b09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2167577 3' similar to contains Alu repetitive element;contains element MIR repetitive element ;
3789	16389	28854	0.93	7.4E-01	AF112538.1	NT	Malva pusilla actin (Act1) mRNA, complete cds
4400	16985	29430	7.7	7.4E-01	AL163246.2	NT	Homo sapiens chromosome 21 segment HS21C046
7785	20328	33234	1.03	7.4E-01	AL161551.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 51
7785	20328	33235	1.03	7.4E-01	AL161551.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 51
8568	21107	34026	0.93	7.4E-01	BF346266.1	EST_HUMAN	602018456F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4154340 5'
8647	21186		0.64	7.4E-01	U87960.1	NT	Rattus norvegicus leukocyte common antigen receptor (LAR) gene, trans-spliced alternative untranslated exon
9026	21563	34492	7.17	7.4E-01	BE747503.1	EST_HUMAN	601573026F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3834174 5'
9083	21619	34554	1.19	7.4E-01	AA187986.1	EST_HUMAN	zp67h01.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone IMAGE:625297 3' similar to SW:TCPQ_MOUSE P42932 T-COMPLEX PROTEIN 1, THETA SUBUNIT ;
10301	22795	35786	0.59	7.4E-01	11424933	NT	Homo sapiens NY-REN-45 antigen (LOC51133), mRNA
11516	23964	37034	1.68	7.4E-01	AB021490.2	NT	Oryzias latipes gene for membrane guanylyl cyclase OlGC1, complete cds
11516	23964	37035	1.68	7.4E-01	AB021490.2	NT	Oryzias latipes gene for membrane guanylyl cyclase OlGC1, complete cds
11677	24096		4.11	7.4E-01	6753217	NT	Mus musculus complement component 1 inhibitor (C1nh), mRNA
11794	24175		1.28	7.4E-01	AI472641.1	EST_HUMAN	ta13h01.x1 NCI_CGAP_Lym5 Homo sapiens cDNA clone IMAGE:2043985 3'
4723	17304	29748	0.72	7.3E-01	AE001166.1	NT	Borrelia burgdorferi (section 52 of 70) of the complete genome
4810	17388	29839	2.93	7.3E-01	AF225421.1	NT	Homo sapiens HT017 mRNA, complete cds
5260	17823	30248	0.99	7.3E-01	O43103	SWISSPROT	FERRICHROME SIDEROPHORE PEPTIDE SYNTHETASE
6720	19314	32116	5.86	7.3E-01	L35772.1	NT	Mus musculus antigen (CD72) gene
6720	19314	32117	5.86	7.3E-01	L35772.1	NT	Mus musculus antigen (CD72) gene
7151	24777	32525	0.82	7.3E-01	AJ011418.1	NT	Lycopersicon esculentum mRNA for ubiquitin activating enzyme
7549	20068	32942	7.77	7.3E-01	M26511.1	NT	V.alginolyticus sucrase (scrB) gene, complete cds
7549	20068	32943	7.77	7.3E-01	M26511.1	NT	V.alginolyticus sucrase (scrB) gene, complete cds
11307	23800	36859	3.86	7.3E-01	AA678019.1	EST_HUMAN	zt25b08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431799 3'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11307	23800	36860	3.86	7.3E-01	AA678019.1	EST_HUMAN	zi25b08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431799 3'
864	13479		1.66	7.2E-01	L29281.1	NT	Rattus norvegicus Initiation factor-2 kinase (eIF-2a) mRNA, complete cds
2000	14582	27141	3.04	7.2E-01	X79140.1	NT	N.tabacum NeIF-4A13 mRNA
2501	15065	27639	1.36	7.2E-01	AB009605.1	NT	Gallus gallus gene for melanocortin 2-receptor, complete cds
3103	15718	28188	1.29	7.2E-01	AF198100.1	NT	Fowlpox virus, complete genome
3500	16105	28580	2.97	7.2E-01	AF065606.1	NT	Giardia intestinalis variant-specific surface protein (vsp417-6) gene, vsp417-6/A-I allele, complete cds
3940	16538	29005	1.81	7.2E-01	BF338350.1	EST_HUMAN	602035589F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183222 5'
4185	16775	29222	0.6	7.2E-01	U02568.1	NT	Dictyocaulus viviparus nematode polyprotein antigen precursor (DvA) mRNA, complete cds
4884	17459	29911	2.54	7.2E-01	D90314.1	NT	L.mesenteroides gene for sucrose phosphorylase (EC 2.4.1.7)
5348	17908	30323	0.9	7.2E-01	AF158800.2	NT	Streptococcus thermophilus bacteriophage Sfi11, complete genome
5386	17945	30358	0.59	7.2E-01	AL161563.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 63
7265	19793	32649	0.82	7.2E-01	U69633.1	NT	Solanum tuberosum cold-stress inducible protein (C17) gene, complete cds
8391	20931	33851	1.15	7.2E-01	AF236061.1	NT	Oryctolagus cuniculus RING-finger binding protein mRNA, partial cds
8893	21431		0.53	7.2E-01	AV743773.1	EST_HUMAN	AV743773 CB Homo sapiens cDNA clone CBMAFD06 5'
10243	22738	35729	2.14	7.2E-01	BF670061.1	EST_HUMAN	602118381F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4275381 5'
10618	23150	36162	5.23	7.2E-01	U82623.1	NT	Rattus norvegicus cytocentrin mRNA, complete cds
12037	16775	29222	1.66	7.2E-01	U02568.1	NT	Dictyocaulus viviparus nematode polyprotein antigen precursor (DvA) mRNA, complete cds
12233	24449		4.42	7.2E-01	AP000063.1	NT	Aeropyrum pernix genomic DNA, section 6/7
12266	24995		1.67	7.2E-01	Y10168.1	NT	B.thuringiensis PK1 & cap genes, putative
721	13341	25831	10.56	7.1E-01	D21070.1	NT	Rana catesbeiana mRNA for bullfrog skeletal muscle calcium release channel (ryanodine receptor) alpha isoform(RyR1), complete cds
3098	15713	28185	18.71	7.1E-01	AJ270777.1	NT	Homo sapiens partial TCF-4 gene for T-cell transcription factor-4, exons 15-16
4287	16873	29321	4.11	7.1E-01	7305360	NT	Mus musculus otogelin (Otog), mRNA
4287	16873	29322	4.11	7.1E-01	7305360	NT	Mus musculus otogelin (Otog), mRNA
6103	18719	31471	1.81	7.1E-01	BF681034.1	EST_HUMAN	602155438F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4296344 5'
6103	18719	31472	1.81	7.1E-01	BF681034.1	EST_HUMAN	602155438F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4296344 5'
7029	19563	32390	6.04	7.1E-01	U36232.1	NT	Drosophila melanogaster 6-pyruvoyltetrahydropterin synthase (pr) gene, complete cds
8132	20673	33584	0.53	7.1E-01	H54244.1	EST_HUMAN	yq89d09.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:202961 3'
8671	21210	34128	0.78	7.1E-01	BE074185.1	EST_HUMAN	RC1-BT0567-301299-011-d09 BT0567 Homo sapiens cDNA
8671	21210	34129	0.78	7.1E-01	BE074185.1	EST_HUMAN	RC1-BT0567-301299-011-d09 BT0567 Homo sapiens cDNA
9769	22267	35252	1.46	7.1E-01	BE904405.1	EST_HUMAN	601496330F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3898495 5'
10309	22803	35795	1.06	7.1E-01	M12961.1	NT	Human T-cell receptor germline gamma-chain J2 gene
12012	24878		2.58	7.1E-01	AA421492.1	EST_HUMAN	zu06h11.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:731109 3'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1272	13868	26387	1.3	7.0E-01	AB014514.1	NT	Homo sapiens mRNA for KIAA0614 protein, partial cds
1272	13868	26388	1.3	7.0E-01	AB014514.1	NT	Homo sapiens mRNA for KIAA0614 protein, partial cds
2492	15057	27630	1.22	7.0E-01	N62412.1	EST_HUMAN	yz73e07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:288708 3' similar to contains Alu repetitive element;
2492	15057	27631	1.22	7.0E-01	N62412.1	EST_HUMAN	yz73e07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:288708 3' similar to contains Alu repetitive element;
5213	17778		1.98	7.0E-01	AL163301.2	NT	Homo sapiens chromosome 21 segment HS21C101
5362	17922	30336	2.99	7.0E-01	AE003921.1	NT	Xylella fastidiosa, section 67 of 229 of the complete genome
8107	18723		1.03	7.0E-01	AB021316.1	NT	Arabidopsis thaliana mRNA for chlorophyll b synthase, complete cds
8319	20860		11.92	7.0E-01	AE000253.1	NT	Escherichia coli K-12 MG1655 section 143 of 400 of the complete genome
9240	21766	34714	0.61	7.0E-01	U53868.1	NT	Clostridium acetobutylicum mannitol-specific phosphotransferase system (PTS) system, mtlA, mtlR, mtlF, and mtlD genes, complete cds
9240	21766	34715	0.61	7.0E-01	U53868.1	NT	Clostridium acetobutylicum mannitol-specific phosphotransferase system (PTS) system, mtlA, mtlR, mtlF, and mtlD genes, complete cds
10999	23513	36546	1.99	7.0E-01	AV763842.1	EST_HUMAN	AV763842 MDS Homo sapiens cDNA clone MDSCHE04 5'
10999	23513	36547	1.99	7.0E-01	AV763842.1	EST_HUMAN	AV763842 MDS Homo sapiens cDNA clone MDSCHE04 5'
12594	24918	30715	1.35	7.0E-01	9630464	NT	Bacteriophage N15 virion, complete genome
1005	13616	26130	10.2	6.9E-01	U69674.1	NT	Candida albicans squalene epoxidase (CAERG1) gene, complete cds and translational regulator gene, partial cds
1005	13616	26131	10.2	6.9E-01	U69674.1	NT	Candida albicans squalene epoxidase (CAERG1) gene, complete cds and translational regulator gene, partial cds
1353	13948	26472	2.8	6.9E-01	AA593530.1	EST_HUMAN	nn28a09.s1 NCI_CGAP_Gas1 Homo sapiens cDNA clone IMAGE:1085176 3'
3256	15868	28348	1.7	6.9E-01	AE002271.2	NT	Chlamydia muridarum, section 3 of 85 of the complete genome
5954	18576	31310	0.8	6.9E-01	AB035662.1	NT	Branchiostoma belcheri BbNA3 mRNA for notochord actin, complete cds
6508	19108	31893	1.31	6.9E-01	BE296188.1	EST_HUMAN	601177333F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3532328 5'
7921	20463	33369	3.4	6.9E-01	AL161573.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 69
7921	20463	33370	3.4	6.9E-01	AL161573.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 69
9098	21634		0.83	6.9E-01	AF118046.1	NT	Entamoeba dispar cation transporting ATPase (atpase) gene, partial cds
9611	22111	35073	0.62	6.9E-01	AF206319.1	NT	Musa acuminata pectate lyase 1 (PL1) mRNA, complete cds
9611	22111	35074	0.62	6.9E-01	AF206319.1	NT	Musa acuminata pectate lyase 1 (PL1) mRNA, complete cds
10307	22801	35793	0.66	6.9E-01	BF242367.1	EST_HUMAN	601880580F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4109419 5'
11138	23646	36687	1.94	6.9E-01	D89013.1	NT	Homo sapiens DAN gene, complete cds
11138	23646	36688	1.94	6.9E-01	D89013.1	NT	Homo sapiens DAN gene, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11651	24870		2.36	6.9E-01	Q99958	SWISSPROT	FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FKH-14)
12670	25003	30611	1.33	6.9E-01	AI888312.1	EST_HUMAN	wn31f02.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2447087 3'
992	13604	26118	1.28	6.8E-01	AF017784.1	NT	Giardia intestinalis carbamate kinase gene, complete cds
2698	15255		1.25	6.8E-01	D90917.1	NT	Synechocystis sp. PCC6803 complete genome, 27/27, 3418852-3573470
2856	14249	26783	1.62	6.8E-01	AA854475.1	EST_HUMAN	aj75e05.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone IMAGE:1402256 3' similar to gb:X56411_rna1 ALCOHOL DEHYDROGENASE CLASS II PI CHAIN (HUMAN);
4672	17254	29706	1.45	6.8E-01	J00762.1	NT	Rat(hooded) prolactin gene : exon iii and flanks
9556	22056	35017	2.11	6.8E-01	AB037766.1	NT	Homo sapiens mRNA for KIAA1345 protein, partial cds
10261	22756		0.48	6.8E-01	AA687936.1	EST_HUMAN	nv13e07.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1220100 3' similar to gb:X13546_rna1 Human HMG-17 gene for non-histone chromosomal protein (HUMAN);
10965	23480	36505	2.96	6.8E-01	AJ276675.1	NT	Stagonospora avenae bgl1 gene for beta-glucosidase, exons 1-4
10965	23480	36506	2.96	6.8E-01	AJ276675.1	NT	Stagonospora avenae bgl1 gene for beta-glucosidase, exons 1-4
10993	23507	36540	2.16	6.8E-01	AF038939.1	NT	Mus musculus zinc finger protein (Peg3) mRNA, complete cds
10993	23507	36541	2.16	6.8E-01	AF038939.1	NT	Mus musculus zinc finger protein (Peg3) mRNA, complete cds
11178	23684	36730	2.2	6.8E-01	AF164151.1	NT	Anopheles gambiae strain M2 translation initiation factor 4C (1A) (eIF-4C) mRNA, complete cds
11475	23925	36995	1.77	6.8E-01	AF110520.1	NT	Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial>
11475	23925	36996	1.77	6.8E-01	AF110520.1	NT	Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial>
320	12974	25463	27.63	6.7E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
361	13010	25493	26.51	6.7E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
1955	14539		0.97	6.7E-01	M12132.1	NT	Quail fast skeletal muscle troponin I gene, complete cds
2192	14768	27340	1.65	6.7E-01	AA451864.1	EST_HUMAN	zx12g12.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:786310 3' similar to contains element TAR1 repetitive element ;
2211	15460	27361	2.66	6.7E-01	AF186073.1	NT	Drosophila melanogaster Mst85C gene, complete cds; NMDMC isoform (Nmdmc) gene, complete cds, alternatively spliced; and transcription factor (Relish) gene, complete cds, alternatively spliced
3026	15642	28120	4.28	6.7E-01	6678580	NT	Mus musculus Wiskott-Aldrich syndrome protein (Wasp), mRNA
4550	17133	29581	0.64	6.7E-01	X74421.1	NT	S.tuberosum mRNA for glucose-6-phosphate dehydrogenase

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO.	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Accession No.	Top Hit Database Source	Top Hit Descriptor
5100	17672	30111	0.96	6.7E-01	AW079110.1	EST_HUMAN	xa95g12.x1 NCI_CGAP_Co17 Homo sapiens cDNA clone IMAGE:2574598 3'
5700	18326	30829	0.8	6.7E-01	J04836.1	NT	M.barkeri ATPase alpha and beta subunit (atpA and atpB) genes, complete cds
5700	18326	30830	0.8	6.7E-01	J04836.1	NT	M.barkeri ATPase alpha and beta subunit (atpA and atpB) genes, complete cds
6116	18732	31485	0.83	6.7E-01	AE001486.1	NT	Helicobacter pylori, strain J99 section 47 of 132 of the complete genome
6465	19066	31851	1.55	6.7E-01	9635035	NT	Gallid herpesvirus 2, complete genome
6465	19066	31852	1.55	6.7E-01	9635035	NT	Gallid herpesvirus 2, complete genome
7356	19882		4.12	6.7E-01	AE004606.1	NT	Pseudomonas aeruginosa PA01, section 167 of 529 of the complete genome
7378	19904	32768	0.9	6.7E-01	AE001486.1	NT	Helicobacter pylori, strain J99 section 47 of 132 of the complete genome
10049	22544		0.87	6.7E-01	M34046.1	NT	Human placental protein 14 (PP14) gene, complete cds
10832	23353	36368	2.52	6.7E-01	BF354649.1	EST_HUMAN	CM3-HT0769-010600-197-c03 HT0769 Homo sapiens cDNA
11333	23031	36040	3.45	6.7E-01	O14357	SWISSPROT	N-ACETYLGLUCOSAMINYL-PHOSPHATIDYLINOSITOL BIOSYNTHETIC PROTEIN GPI1
2548	15110	27682	2.66	6.6E-01	AF075240.1	NT	Homo sapiens SLIT1 protein (SLIL2) mRNA, partial cds
2724	15279	27846	1.01	6.6E-01	AF199339.1	NT	Homo sapiens lens epithelium-derived growth factor gene, alternatively spliced, complete cds
3536	16141	28623	1.35	6.6E-01	4506880	NT	Homo sapiens sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A (SEMA5A) mRNA
3719	16320	28788	3.42	6.6E-01	Y07669.1	NT	C.albicans random DNA marker, 282bp
4187	16777		0.67	6.6E-01	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
5227	17791	30210	0.97	6.6E-01	AI218230.1	EST_HUMAN	qh23a10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1845498 3' similar to contains PTR5.b2 MER28 repetitive element ;
6474	19075	31858	4.22	6.6E-01	6680577	NT	Mus musculus kinesin light chain 2 (Klc2), mRNA
7675	20186	33074	3.61	6.6E-01	AV660506.1	EST_HUMAN	AV660506 GLC Homo sapiens cDNA clone GLCGID04 3'
8501	21040	33961	0.84	6.6E-01	AV704700.1	EST_HUMAN	AV704700 ADB Homo sapiens cDNA clone ADBCAF11 5'
9582	22082		1.73	6.6E-01	AL163278.2	NT	Homo sapiens chromosome 21 segment HS21C078
9915	22411		0.66	6.6E-01	AU118198.1	EST_HUMAN	AU118198 HEMBA1 Homo sapiens cDNA clone HEMBA1003079 5'
12118	24377	30973	1.27	6.6E-01	AF110001.1	NT	Homo sapiens guanylate cyclase activating protein 3 (GCAP3) gene, exon 4 and complete cds
651	13274	25751	1.12	6.5E-01	M75140.1	NT	H.vulgaris Na,K-ATPase alpha subunit mRNA, complete cds
651	13274	25752	1.12	6.5E-01	M75140.1	NT	H.vulgaris Na,K-ATPase alpha subunit mRNA, complete cds
3480	16086	28560	5.04	6.5E-01	AB041225.1	NT	Mus musculus gene for Tob2, complete cds
4110	16704	29157	1.1	6.5E-01	4504632	NT	Homo sapiens interleukin 10 receptor, alpha (IL10RA) mRNA
4369	16956	29398	3.29	6.5E-01	AJ272265.1	NT	Homo sapiens SPP2 gene for secreted phosphoprotein 24 precursor, exons 1-8
4699	17281	29728	1.28	6.5E-01	D00584.1	NT	Oryza sativa gene for prepro-glutelin, exons 1, 2, 3, 4, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5219	17784	30202	2.39	6.5E-01	U28921.1	NT	Phaseolus vulgaris ATPase gamma subunit mRNA, nuclear gene encoding mitochondrial protein, partial cds
5336	17897	30312	1.02	6.5E-01	Z70628.1	NT	H.sapiens mRNA for immunoglobulin heavy chain variable region (9D4-A6, VH4, 4-59/DP-71)
6825	19415	32231	1.26	6.5E-01	D88348.1	NT	Chicken mRNA for 115-kDa melanosomal matrix protein, complete cds
7663	20175	33062	0.96	6.5E-01	AI799882.1	EST_HUMAN	wc46a02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321642 3'
9751	22249		1.25	6.5E-01	T78904.1	EST_HUMAN	yd21b04.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:108847 3'
10238	22733	35725	2.49	6.5E-01	AF119676.1	NT	Mus musculus small GTP-binding protein RAB25 (Rab25) gene, complete cds
10512	23050	36061	3.35	6.5E-01	H87583.1	EST_HUMAN	yw17f06.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:252515 5'
10566	23102	36116	4.35	6.5E-01	AA601287.1	EST_HUMAN	no15c07.s1 NCI_CGAP_Pne1 Homo sapiens cDNA clone IMAGE:1100748 3'
10669	23201		4.29	6.5E-01	AU138078.1	EST_HUMAN	AU138078 PLACE1 Homo sapiens cDNA clone PLACE1007810 5'
11470	23920	36989	2.7	6.5E-01	AF014115.1	NT	Plasmodium berghei cytochrome c oxidase subunit III, cytochrome c oxidase subunit I, and cytochrome b genes, mitochondrial genes encoding mitochondrial proteins, complete cds
12067	24348		8.24	6.5E-01	BE465050.1	EST_HUMAN	hv74a10.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3179130 3'
12321	24617		3.04	6.5E-01	Z74145.1	NT	S.cerevisiae chromosome IV reading frame ORF YDL097c
273	12930	25417	9.34	6.4E-01	U48848.1	NT	Drosophila melanogaster 8kd dynein light chain mRNA, complete cds
3502	16107	28583	3.78	6.4E-01	U48854.2	NT	Mus musculus dystroglycan 1 (DAG1) gene, exons 1 and 2 and complete cds
3928	16526	28993	1.33	6.4E-01	AB046827.1	NT	Homo sapiens mRNA for KIAA1607 protein, partial cds
4591	17174	29619	0.66	6.4E-01	Y12488.1	NT	M.musculus whn gene
4591	17174	29620	0.66	6.4E-01	Y12488.1	NT	M.musculus whn gene
5402	17960	30371	0.97	6.4E-01	AE002551.2	NT	Neisseria meningitidis serogroup B strain MC58 section 193 of 206 of the complete genome
8549	21088	34010	1.76	6.4E-01	AE001247.1	NT	Treponema pallidum section 83 of 87 of the complete genome
10001	22496	35486	8.26	6.4E-01	U82828.1	NT	Homo sapiens ataxia telangiectasia (ATM) gene, complete cds
10015	22510	35501	1.16	6.4E-01	BF670405.1	EST_HUMAN	602150289F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4291126 5'
12188	24420		29.97	6.4E-01	AV759212.1	EST_HUMAN	AV759212 MDS Homo sapiens cDNA clone MDSCGC09 5'
459	13093	25587	3.75	6.3E-01	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
560	13191	25669	56.3	6.3E-01	U32689.1	NT	Haemophilus influenzae Rd section 4 of 163 of the complete genome
2207	14783	27356	3.24	6.3E-01	U81136.1	NT	Shigella flexneri multi-antibiotic resistance locus
2614	15176	27744	2.78	6.3E-01	U75331.1	NT	Gallus gallus bone morphogenetic protein 1 (BMP1) mRNA, partial cds
2614	15176	27745	2.78	6.3E-01	U75331.1	NT	Gallus gallus bone morphogenetic protein 1 (BMP1) mRNA, partial cds
3050	15666		0.75	6.3E-01	Y17275.1	NT	Lycopersicon esculentum p69a gene, complete CDS
6214	18824	31595	0.78	6.3E-01	BE093906.1	EST_HUMAN	PM0-BT0757-010500-002-a05 BT0757 Homo sapiens cDNA
6712	19306	32110	1	6.3E-01	L27798.1	NT	Streptococcus dysgalactiae (mag) gene, complete cds
6712	19306	32111	1	6.3E-01	L27798.1	NT	Streptococcus dysgalactiae (mag) gene, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8458	20998		3.32	6.3E-01	BE902044.1	EST_HUMAN	601676889F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959351 5'
8819	21358	34284	0.91	6.3E-01	S62927.1	NT	glycoprotein IIIa (Alu 1 and 3 fusion junction) [human, Genomic Mutant, 300 nt]
9147	21682	34627	1.15	6.3E-01	BF216984.1	EST_HUMAN	601884050F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4102586 5'
9341	21855	34804	2.9	6.3E-01	9627521	NT	Variola virus, complete genome
9341	21855	34805	2.9	6.3E-01	9627521	NT	Variola virus, complete genome
9851	22349		0.67	6.3E-01	AE002329.2	NT	Chlamydia muridarum, section 59 of 85 of the complete genome
10324	22818	35814	1.52	6.3E-01	Z73003.1	NT	S.cerevisiae chromosome VII reading frame ORF YGR218w
10421	22915	35915	0.87	6.3E-01	AE000313.1	NT	Escherichia coli K-12 MG1655 section 203 of 400 of the complete genome
10939	23456	36479	2.45	6.3E-01	AA877715.1	EST_HUMAN	nr09h06.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1161371 3' similar to TR:O02916 O02916 HLARK. ;
11216	23719	36773	15.21	6.3E-01	AI904160.1	EST_HUMAN	CM-BT043-090299-046 BT043 Homo sapiens cDNA
11302	23795	36853	1.94	6.3E-01	P47003	SWISSPROT	HYPOTHETICAL 13.7 KD PROTEIN IN INO1-IDS2 INTERGENIC REGION
11458	23908	36975	2.02	6.3E-01	P36073	SWISSPROT	HYPOTHETICAL 15.3 KD PROTEIN IN VMA12-APN1 INTERGENIC REGION
11769	25042	30505	30.63	6.3E-01	9910293	NT	Mus musculus keratin complex 2, gene 6g (Krt2-6g), mRNA
11864	24219		1.85	6.3E-01	AF105227.1	NT	Homo sapiens 3'-phosphoadenosine 5'-phosphosulfate synthetase (PAPSS) mRNA, complete cds
12082	24953		3.2	6.3E-01	X83528.1	NT	C.limicola pscD gene
5175	17742	30171	0.71	6.2E-01	AF157898.1	NT	Spermophilus susilicus isolate S47 cytochrome b (cytb) gene, complete cds; mitochondrial gene for mitochondrial product
6030	18649	31390	2.03	6.2E-01	Q10135	SWISSPROT	HYPOTHETICAL 142.5 KD PROTEIN C23E2.02 IN CHROMOSOME I
7506	20028		3.14	6.2E-01	AF022253.1	NT	Mus musculus calcium-sensing receptor related protein 4 (Casr-rs4) mRNA, partial cds
7548	24786	32941	1.08	6.2E-01	AL021127.2	NT	Mus musculus chromosome X contigA; putative Magea9 gene, Caltractin, NAD(P) steroid dehydrogenase and Zinc finger protein 185
8243	20784	33703	5.65	6.2E-01	H72255.1	EST_HUMAN	ys01e08.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:213542 3'
8790	21329	34254	0.54	6.2E-01	AF034411.1	NT	Lycopersicon esculentum cytosolic Cu,Zn superoxide dismutase (Sod) gene, partial cds; and dehydroquinate dehydratase/shikimate:NADP oxidoreductase gene, complete cds
9370	20309	33212	1.75	6.2E-01	BE562687.1	EST_HUMAN	601336146F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3690010 5'
9429	21938		2.35	6.2E-01	M24461.1	NT	Human pulmonary surfactant-associated protein SP-B (SFTP3) mRNA, complete cds
9990	22486	35472	5.85	6.2E-01	AL161511.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 23
10429	22923	35927	3.76	6.2E-01	P27410	SWISSPROT	NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE ; THIOL PROTEASE P3C ; HELICASE (2C LIKE PROTEIN); COAT PROTEIN]
10429	22923	35928	3.76	6.2E-01	P27410	SWISSPROT	NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE ; THIOL PROTEASE P3C ; HELICASE (2C LIKE PROTEIN); COAT PROTEIN]
2438	15005		4.95	6.1E-01	6678076	NT	Mus musculus secreted acidic cysteine rich glycoprotein (Sparc), mRNA
4632	17215	29666	1.05	6.1E-01	4557538	NT	Homo sapiens solute carrier family 26 (sulfate transporter), member 2 (SLC26A2) mRNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5141	17712	30142	1.09	6.1E-01	L20427.1	NT	Rattus norvegicus dihydroxypolyprenylbenzoate methyltransferase mRNA, complete cds
5141	17712	30143	1.09	6.1E-01	L20427.1	NT	Rattus norvegicus dihydroxypolyprenylbenzoate methyltransferase mRNA, complete cds
5727	18353	31057	1.54	6.1E-01	M59940.1	NT	Caenorhabditis elegans N2 CeMyoD (hlh-1) alternatively spliced genes, complete cds
6951	19528	32351	3.55	6.1E-01	M64733.1	NT	Rat TRPM-2 gene, complete cds
6951	19528	32352	3.55	6.1E-01	M64733.1	NT	Rat TRPM-2 gene, complete cds
8175	20716	33632	3.57	6.1E-01	AF033535.1	NT	Arabidopsis thaliana putative zinc transporter (ZIP1) mRNA, complete cds
8730	21269	34187	1.23	6.1E-01	11431065	NT	Homo sapiens mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4), mRNA
8730	21269	34188	1.23	6.1E-01	11431065	NT	Homo sapiens mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4), mRNA
9336	21850	34798	19.4	6.1E-01	AF236117.1	NT	Homo sapiens G-protein coupled receptor EDG-7 mRNA, complete cds
9336	21850	34799	19.4	6.1E-01	AF236117.1	NT	Homo sapiens G-protein coupled receptor EDG-7 mRNA, complete cds
9756	22254	35236	1.15	6.1E-01	AE004452.1	NT	Pseudomonas aeruginosa PA01, section 13 of 529 of the complete genome
9959	22454	35436	1.8	6.1E-01	AF119117.1	NT	Homo sapiens dopamine transporter (SLC6A3) gene, complete cds
10837	23358		8.53	6.1E-01	X74507.1	NT	P.sativum mdh mRNA for chloroplast malate dehydrogenase (NADP+)
11581	24027	37095	2.19	6.1E-01	S83182.1	NT	hyaluronan-binding protein=hepatocyte growth factor activator homolog [human, plasma, mRNA, 2408 nt]
11581	24027	37096	2.19	6.1E-01	S83182.1	NT	hyaluronan-binding protein=hepatocyte growth factor activator homolog [human, plasma, mRNA, 2408 nt]
12530	24643		1.91	6.1E-01	X95287.1	NT	M.mazei orfA, orfB, and orfC of archaeal ABC-transporter system
520	13152	25635	1.46	6.0E-01	D87675.1	NT	Homo sapiens DNA for amyloid precursor protein, complete cds
587	13217		3.41	6.0E-01	5802999	NT	Homo sapiens adaptor-related protein complex 3, mu 2 subunit (CLA20), mRNA
1406	13999	26528	1.93	6.0E-01	AF065253.1	NT	Human respiratory syncytial virus strain CH93-53b attachment protein (G) gene, complete cds
3887	16485	28946	0.86	6.0E-01	AJ233396.1	NT	Viral hemorrhagic septicemia virus N, P, M, G, Nv, L genes, French strain 07-71
4267	16853		1.16	6.0E-01	AF056895.1	NT	Homo sapiens Notch3 (NOTCH3) gene, exons 26, 27, and 28
5485	18119	30526	1.93	6.0E-01	P20288	SWISSPROT	D(2) DOPAMINE RECEPTOR
5631	18260	30732	2.28	6.0E-01	AW139713.1	EST_HUMAN	UI-H-BI1-aeb-a-10-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:2718619 3'
6660	19256	32059	3.73	6.0E-01	U38813.1	NT	Musca domestica insecticide-susceptible strain voltage-sensitive sodium channel mRNA, complete cds
6767	19360	32169	0.79	6.0E-01	Q04912	SWISSPROT	MACROPHAGE-STIMULATING PROTEIN RECEPTOR PRECURSOR (MSP RECEPTOR) (P185-RON) (CDW136) (CD136 ANTIGEN)
7391	19916	32780	5.29	6.0E-01	AJ277661.1	NT	Homo sapiens partial LMO1 gene for LIM domain only 1 protein, exon 1
8066	20608	33520	4.72	6.0E-01	P02835	SWISSPROT	SEGMENTATION PROTEIN FUSHI TARAZU
8066	20608	33521	4.72	6.0E-01	P02835	SWISSPROT	SEGMENTATION PROTEIN FUSHI TARAZU
9737	22235	35214	2.22	6.0E-01	AB008193.1	NT	Homo sapiens genes for leukotriene B4 receptor BLT2, leukotriene B4 receptor BLT1, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10174	22689		1.61	6.0E-01	Q01497	SWISSPROT	PEROXISOMAL MEMBRANE PROTEIN PER9 (PEROXIN-3)
10936	23453	36476	2.14	6.0E-01	AJ131892.1	NT	Gallus gallus mRNA for Hyperion protein, 419 kD isoform
10936	23453	36477	2.14	6.0E-01	AJ131892.1	NT	Gallus gallus mRNA for Hyperion protein, 419 kD isoform
11426	23877	36942	2.84	6.0E-01	AI420623.1	EST_HUMAN	tf08f07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2095621 3'
12158	24398	30978	1.82	6.0E-01	11421663	NT	Homo sapiens nuclear factor (erythroid-derived 2)-like 3 (NFE2L3), mRNA
12265	24475		1.99	6.0E-01	AA706087.1	EST_HUMAN	zj96g05.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:462776 3'
12426	24879		1.29	6.0E-01	5803136	NT	Homo sapiens RNA binding motif protein 3 (RBM3), mRNA
12469	24885	30709	2.49	6.0E-01	9055303	NT	Mus musculus cGMP-inhibited phosphodiesterase (Pde3a), mRNA
12499	24810		6.82	6.0E-01	BE157617.1	EST_HUMAN	RC1-HT0375-030500-015-c03 HT0375 Homo sapiens cDNA
1038	13648	26160	1.09	5.9E-01	U32701.1	NT	Haemophilus influenzae Rd section 16 of 163 of the complete genome
1447	14039	26568	1.06	5.9E-01	6680232	NT	Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A lyase (Hmgcl), mRNA
3308	15919	28395	5.12	5.9E-01	AL163267.2	NT	Homo sapiens chromosome 21 segment HS21C067
3308	15919	28396	5.12	5.9E-01	AL163267.2	NT	Homo sapiens chromosome 21 segment HS21C067
4304	16890		4.32	5.9E-01	AF162756.1	NT	Rattus norvegicus cenexin 2 mRNA, partial cds
6591	19188	31991	1.48	5.9E-01	AF065440.2	NT	Homo sapiens low density lipoprotein receptor-related protein II (LRP2) gene, exon 1 and partial cds
7310	19838	32696	5.58	5.9E-01	AB023486.1	NT	Homo sapiens gene for histamine H2 receptor, promoter region and complete cds
7941	20483	33395	0.57	5.9E-01	D90911.1	NT	Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643
9462	21987	34943	0.93	5.9E-01	AF063204.2	NT	Chlamydia trachomatis strain K/UW31/Cx major outer membrane protein (omp1) gene, complete cds
9827	22325		0.68	5.9E-01	P06463	SWISSPROT	E6 PROTEIN
10091	22586	35579	1.15	5.9E-01	P55284	SWISSPROT	VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
10551	23087	36102	3.24	5.9E-01	Q9X0I3	SWISSPROT	THYMIDYLATE KINASE (DTMP KINASE)
10557	23093	36105	1.75	5.9E-01	AF197944.1	NT	Xenopus laevis receptor protein tyrosine phosphatase delta (XPTP-D) mRNA, complete cds
10840	23361	36376	3	5.9E-01	AW937175.1	EST_HUMAN	PM1-DT0041-190100-002-h03 DT0041 Homo sapiens cDNA
11073	23585	36626	2.25	5.9E-01	AF064626.1	NT	Mus spretus strain SPRET/Ei CD48 antigen (Cd48) gene, partial cds
11810	24182	31030	1.92	5.9E-01	L42320.1	NT	Oryctolagus cuniculus alpha 1 anti-trypsin (alpha 1 AT) gene, promoter region
12053	24336		2.88	5.9E-01	AB017705.1	NT	Aspergillus oryzae pyrG gene for orotidine-5'-phosphate decarboxylase, complete cds
12280	24483		7.56	5.9E-01	P34926	SWISSPROT	MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2]
1952	14536	27092	1.8	5.8E-01	P40472	SWISSPROT	SIM1 PROTEIN
4056	16653	29119	1.22	5.8E-01	BF695736.1	EST_HUMAN	601852474F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4076131 5'
4612	17195	29641	3.73	5.8E-01	AB009077.1	NT	Vigna radiata mRNA for proton pyrophosphatase, complete cds
4914	17489		1.18	5.8E-01	AF110848.1	NT	Megaselia scalaris sex-lethal homolog (Megsxl) gene, partial cds, alternatively spliced products
5577	18208		0.75	5.8E-01	AE002152.1	NT	Ureaplasma urealyticum section 53 of 59 of the complete genome

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5722	18348	31051	2.52	5.8E-01	Q10699	SWISSPROT	POTENTIAL 5'-3' EXONUCLEASE
6331	18937	31713	2.37	5.8E-01	D78659.1	EST_HUMAN	HUM500E06B Human placenta polyA+ (TFujiwara) Homo sapiens cDNA clone GEN-500E06 5'
6454	19055	31840	0.71	5.8E-01	D50601.1	NT	Shigella sonnei DNA for 26 ORFs, complete cds
6903	19637		2.47	5.8E-01	S65091.1	NT	cyclic AMP-regulated phosphoprotein [rats, mRNA, 1030 nt]
7828	20370		2.57	5.8E-01	H41571.1	EST_HUMAN	yn91b03.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:175757 3' similar to gb:S78187 M-PHASE INDUCER PHOSPHATASE 2 (HUMAN);
8031	20573	33477	0.66	5.8E-01	AI280051.1	EST_HUMAN	qh85d10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1853779 3'
8031	20573	33478	0.66	5.8E-01	AI280051.1	EST_HUMAN	qh85d10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1853779 3'
8131	20672	33582	2.34	5.8E-01	P14328	SWISSPROT	SPORE COAT PROTEIN SP96
8131	20672	33583	2.34	5.8E-01	P14328	SWISSPROT	SPORE COAT PROTEIN SP96
8823	21362	34287	9.48	5.8E-01	AJ270774.1	NT	Homo sapiens partial TCF-4 gene for T-cell transcription factor-4, exons 6-11
8902	21440	34363	0.88	5.8E-01	Q27368	SWISSPROT	TRANSCRIPTION FACTOR E2F
8903	21441	34364	0.56	5.8E-01	Q20471	SWISSPROT	PUTATIVE CASEIN KINASE I F46F2.2 IN CHROMOSOME X
9514	22014		0.89	5.8E-01	BF031606.1	EST_HUMAN	601557774F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3827298 5'
10869	23390	36405	9.44	5.8E-01	AJ243213.1	NT	Homo sapiens partial 5-HT4 receptor gene, exons 2 to 5
10915	23434		3.66	5.8E-01	BF700092.1	EST_HUMAN	602127577F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4284403 5'
11021	23535		2.04	5.8E-01	BF700092.1	EST_HUMAN	602127577F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4284403 5'
3079	15694		0.66	5.7E-01	6755253	NT	Mus musculus plasmacytoma variant translocation 1 (Pvt1), mRNA
3260	15872	28352	1.58	5.7E-01	Q9WTJ2	SWISSPROT	PUTATIVE TRANSCRIPTION FACTOR OVO-LIKE 1 (MOVO1) (MOVO1A)
3552	16156		2.63	5.7E-01	AB033503.1	NT	Populus euramericana peacs-2 mRNA for 1-aminocyclopropane-1-carboxylate synthase, complete cds
3973	16571	29041	3.09	5.7E-01	AF011581.1	NT	Homo sapiens T cell receptor beta chain (BV6S7*2-BJ1S1) mRNA, partial cds
6496	19097	31881	3.67	5.7E-01	BF035413.1	EST_HUMAN	601454962F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3858590 5'
6812	19403	32219	0.72	5.7E-01	AA194201.1	EST_HUMAN	zr38c06.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:685674 5'
6945	18053	30476	1.28	5.7E-01	AL111440.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
7741	20249	33142	1.97	5.7E-01	P00373	SWISSPROT	PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE)
7911	20453		0.57	5.7E-01	AJ251835.1	NT	Mus musculus Kcnq1, Ltrpc5, Mash2, Tape-1, Tssc4 and Tssc6 genes, alternative transcripts
9715	22213	35186	1.17	5.7E-01	AL161532.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 32
9715	22213	35187	1.17	5.7E-01	AL161532.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 32
10468	22962	35973	0.86	5.7E-01	BF540962.1	EST_HUMAN	602067712F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4066610 5'
3410	16019	28498	1	5.6E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
3410	16019	28499	1	5.6E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
4324	16910	29351	0.69	5.6E-01	D83135.1	NT	Chicken TBP gene, exon8, complete cds
8738	21277	34200	4.42	5.6E-01	AV684703.1	EST_HUMAN	AV684703 GKC Homo sapiens cDNA clone GKCFSF05 5'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8738	21277	34201	4.42	5.6E-01	AV684703.1	EST_HUMAN	AV684703 GKC Homo sapiens cDNA clone GKCFSF05 5'
9297	21897	34844	1.11	5.6E-01	AB038782.1	NT	Homo sapiens MUC3A gene for intestinal mucin, partial cds
11658	24085		2.5	5.6E-01	BE888280.1	EST_HUMAN	601514007F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3915457 5'
11779	24166	36775	1.28	5.6E-01	AA493535.1	EST_HUMAN	ng75g10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:940674 similar to contains element PTR7 repetitive element ;
12156	18028	30490	3.31	5.6E-01	AL161501.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 13
12185	24419		2.56	5.6E-01	P50505	SWISSPROT	HIGH AFFINITY POTASSIUM TRANSPORTER
12619	24698		3.11	5.6E-01	BF573829.1	EST_HUMAN	602132029F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271334 5'
1253	13850	26367	1.13	5.5E-01	8393912	NT	Rattus norvegicus Propionyl Coenzyme A carboxylase, beta polypeptide (Pccb), mRNA
2725	15280	27847	13.6	5.5E-01	P03341	SWISSPROT	GAG POLYPROTEIN [CONTAINS: INNER COAT PROTEIN P12; CORE PROTEIN P15; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]
2725	15280	27848	13.6	5.5E-01	P03341	SWISSPROT	GAG POLYPROTEIN [CONTAINS: INNER COAT PROTEIN P12; CORE PROTEIN P15; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]
2943	15559	28033	0.69	5.5E-01	5902085	NT	Homo sapiens superkiller viralicidic activity 2 (S. cerevisiae homolog)-like (SKIV2L), mRNA
3102	15717		1.51	5.5E-01	H46219.1	EST_HUMAN	yo18a10.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:178266 3'
3271	15883	28365	2.68	5.5E-01	AF227240.1	NT	Rabbit oral papillomavirus, complete genome
3755	16356	28825	0.97	5.5E-01	P48755	SWISSPROT	FOS-RELATED ANTIGEN-1
8386	20926	33846	0.66	5.5E-01	AI791766.1	EST_HUMAN	or82c01.y5 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1602336 5'
9682	22181		0.74	5.5E-01	U88415.1	NT	Crimean-Congo hemorrhagic fever virus strain SPU 415/85 nucleoprotein gene, complete cds
10279	22774	35763	0.84	5.5E-01	T05047.1	EST_HUMAN	EST02935 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCQ35
151	12814	25301	12.97	5.4E-01	7657266	NT	Homo sapiens KIAA0929 protein Msx2 interacting nuclear target (MINT) homolog (KIAA0929), mRNA
151	12814	25302	12.97	5.4E-01	7657266	NT	Homo sapiens KIAA0929 protein Msx2 interacting nuclear target (MINT) homolog (KIAA0929), mRNA
611	13239	25713	1.6	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
611	13239	25714	1.6	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
1314	13908	26428	2.58	5.4E-01	AW896087.1	EST_HUMAN	QV4-NN0040-070400-160-c04 NN0040 Homo sapiens cDNA
2154	14731		3.6	5.4E-01	AE002247.2	NT	Chlamydophila pneumoniae AR39, section 74 of 94 of the complete genome
2296	14870	27446	2.18	5.4E-01	AJ276682.1	NT	Drosophila melanogaster mRNA for 15,15' beta carotene dioxygenase (beta-diox gene)
3984	16582	29053	0.62	5.4E-01	U07561.1	NT	Human ABL gene, exon 1b and intron 1b, and putative M8604 Met protein (M8604 Met) gene, complete cds
5259	17822		1.04	5.4E-01	AW747972.1	EST_HUMAN	QV0-BT0041-061099-033-e02 BT0041 Homo sapiens cDNA

WO 01/57277

PCT/US01/00699

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5388	13239	25713	0.59	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
5388	13239	25714	0.59	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
5838	18462	31185	0.81	5.4E-01	AW842327.1	EST_HUMAN	PM2-CN0030-030200-003-c10 CN0030 Homo sapiens cDNA
6338	18944	31723	1.49	5.4E-01	AB025017.1	NT	Rattus norvegicus gene for TIS11, complete cds
7094	19685	32504	1.1	5.4E-01	BE966592.2	EST_HUMAN	601660276R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906090 3'
7374	19900	32762	0.75	5.4E-01	Z21619.1	NT	S.cerevisiae RIB3 gene encoding DBP synthase
7374	19900	32763	0.75	5.4E-01	Z21619.1	NT	S.cerevisiae RIB3 gene encoding DBP synthase
7376	19902	32766	1.47	5.4E-01	Q64428	SWISSPROT	MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA) [INCLUDES: LONG-CHAIN ENOYL-COA HYDRATASE ; LONG CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE]
9901	22398		1.98	5.4E-01	BF572536.1	EST_HUMAN	602076545F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4243690 5'
10957	23472	36497	3.25	5.4E-01	P36858	SWISSPROT	NITRATE REDUCTASE [NADPH] (NR)
11485	23934	37004	5.79	5.4E-01	Q60675	SWISSPROT	LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY CHAIN)
11485	23934	37005	5.79	5.4E-01	Q60675	SWISSPROT	LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY CHAIN)
11586	18944	31723	2.42	5.4E-01	AB025017.1	NT	Rattus norvegicus gene for TIS11, complete cds
11725	24132		2.52	5.4E-01	AI858398.1	EST_HUMAN	wl37g04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2427126 3' similar to gb:M13452 LAMIN A (HUMAN);
542	13173	25653	2.29	5.3E-01	AF019413.1	NT	Homo sapiens HLA class III region containing tenascin X (tenascin-X) gene, partial cds; cytochrome P450 21-hydroxylase (CYP21B), complement component C4 (C4B) G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes,>
2811	15363	27931	6.51	5.3E-01	4506328	NT	Homo sapiens protein tyrosine phosphatase, receptor-type, zeta polypeptide 1 (PTPRZ1) mRNA
2811	15363	27932	6.51	5.3E-01	4506328	NT	Homo sapiens protein tyrosine phosphatase, receptor-type, zeta polypeptide 1 (PTPRZ1) mRNA
3280	15891	28370	3.13	5.3E-01	AF087658.1	NT	Homo sapiens secreted C-type lectin precursor (LSLCL) gene, complete cds
4290	16876		1.39	5.3E-01	U39687.1	NT	Mycoplasma genitalium section 9 of 51 of the complete genome
5649	18277	30753	1.91	5.3E-01	AI820921.1	EST_HUMAN	zu42h12.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:740711 5'
5649	18277	30754	1.91	5.3E-01	AI820921.1	EST_HUMAN	zu42h12.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:740711 5'
5742	18368	31075	0.87	5.3E-01	AA193672.1	EST_HUMAN	zr42g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:668112 5'
5742	18368	31076	0.87	5.3E-01	AA193672.1	EST_HUMAN	zr42g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:668112 5'
5827	18451	31174	1.84	5.3E-01	BE645620.1	EST_HUMAN	7e73c12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3288118 3' similar to gb:J02783 PROTEIN DISULFIDE ISOMERASE PRECURSOR (HUMAN);
5827	18451	31175	1.84	5.3E-01	BE645620.1	EST_HUMAN	7e73c12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3288118 3' similar to gb:J02783 PROTEIN DISULFIDE ISOMERASE PRECURSOR (HUMAN);

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8835	21374		1.83	5.3E-01	L01950.2	NT	Roridula gorgonias ribulose 1,5-bisphosphate carboxylase (rbcL) gene, partial cds; chloroplast gene for chloroplast product
8885	21423	34348	0.63	5.3E-01	BF433956.1	EST_HUMAN	7q71c12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE: 3' similar to contains element MER29 repetitive element ;
8885	21423	34349	0.63	5.3E-01	BF433956.1	EST_HUMAN	7q71c12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE: 3' similar to contains element MER29 repetitive element ;
10112	22607	35597	0.48	5.3E-01	AI954210.1	EST_HUMAN	wx94b02.x1 NCI_CGAP_Mel15 Homo sapiens cDNA clone IMAGE:2551275 3' similar to SW:COXA_HUMAN P20674 CYTOCHROME C OXIDASE POLYPEPTIDE VA PRECURSOR ;
11435	23885	36952	6.92	5.3E-01	BE566291.1	EST_HUMAN	601339867F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3682168 5'
11650	24881		4.22	5.3E-01	AA916053.1	EST_HUMAN	og30e05.s1 NCI_CGAP_Br7 Homo sapiens cDNA clone IMAGE:1441376 3' similar to gb:J02611 APOLIPOPROTEIN D PRECURSOR (HUMAN);
849	13465	25973	19.16	5.2E-01	L20770.1	NT	Drosophila melanogaster helix-loop-helix mRNA, complete cds
1206	13806	26319	10.07	5.2E-01	Q9WV30	SWISSPROT	NUCLEAR FACTOR OF ACTIVATED T CELLS 5 (T CELL TRANSCRIPTION FACTOR NFAT5) (NF-AT5) (REL DOMAIN-CONTAINING TRANSCRIPTION FACTOR NFAT5)
1233	13832	26346	2.91	5.2E-01	AF224492.1	NT	Homo sapiens phospholipid scramblase 1 gene, complete cds
1930	14514		4.11	5.2E-01	AL163285.2	NT	Homo sapiens chromosome 21 segment HS21C085
2191	14767	27339	2.97	5.2E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
3153	15767	28233	1.67	5.2E-01	U65942.1	NT	Chlamydophila abortus strain S26/3 POMP91A and POMP90A precursor, genes, complete cds
3274	15886		0.71	5.2E-01	D73443.1	NT	Azotobacter vinelandii icd gene for isocitrate dehydrogenase, complete cds
3452	16059		1.74	5.2E-01	AL116780.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
3492	16097	28572	2.49	5.2E-01	AA984165.1	EST_HUMAN	am77g05.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:1616504 3'
3694	16295		0.92	5.2E-01	AF020269.1	NT	Medicago sativa chloroplast malate dehydrogenase precursor (p1mdh) mRNA, nuclear gene encoding chloroplast protein, complete cds
5161	17730		0.87	5.2E-01	7106444	NT	Mus musculus vanilloid receptor-like protein 1 (Vrl1), mRNA
5314	17876		0.99	5.2E-01	AL163281.2	NT	Homo sapiens chromosome 21 segment HS21C081
5834	18458	31179	0.97	5.2E-01	AA284261.1	EST_HUMAN	zc44d09.T7 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:325169 3'
9646	24795	35115	1.19	5.2E-01	X02218.1	NT	Chicken duplicated genes for histone H2A, H4 and a histone H3 gene
9646	24795	35116	1.19	5.2E-01	X02218.1	NT	Chicken duplicated genes for histone H2A, H4 and a histone H3 gene
9845	22343	35325	0.64	5.2E-01	AA194518.1	EST_HUMAN	zq05b09.r1 Stratagene muscle 937209 Homo sapiens cDNA clone IMAGE:628793 5'
9940	22435	35411	1.65	5.2E-01	AF143952.2	NT	Homo sapiens PELOTA (PELOTA) gene, complete cds
12590	24682		4.94	5.2E-01	P18516	SWISSPROT	RETINOIC ACID RECEPTOR GAMMA (RAR-GAMMA) (RETINOIC ACID RECEPTOR DELTA) (RAR-DELTA)
645	13268	25746	2.13	5.1E-01	M58509.1	NT	Human adrenodoxin reductase gene, exons 3 to 12
676	13300	25781	3.98	5.1E-01	AJ233944.1	NT	Polyangium vitellinum (strain Pl vt1) 16S rRNA gene

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
676	13300	25782	3.98	5.1E-01	AJ233944.1	NT	Polyangium vitellinum (strain Pl vt1) 16S rRNA gene
1692	14284		0.88	5.1E-01	X87885.1	NT	R.norvegicus mRNA for mammalian fusca protein
2069	14649		11.33	5.1E-01	BF683095.1	EST_HUMAN	602139319F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4298117 5'
4151	16743	29197	4.61	5.1E-01	AI858495.1	EST_HUMAN	wl39b12.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2427263 3'
4266	16852	29300	3.03	5.1E-01	P96380	SWISSPROT	TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)
5229	17793		0.71	5.1E-01	BE091796.1	EST_HUMAN	IL2-BT0731-250400-077-G08 BT0731 Homo sapiens cDNA
6422	19025		0.79	5.1E-01	AV712326.1	EST_HUMAN	AV712326 DCA Homo sapiens cDNA clone DCAAUF07 5'
6997	19495	32316	1.42	5.1E-01	R80873.1	EST_HUMAN	yi94a09.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:146872 3'
8507	21046	33966	0.73	5.1E-01	AW806881.1	EST_HUMAN	QV4-ST0023-160400-172-a01 ST0023 Homo sapiens cDNA
8507	21046	33967	0.73	5.1E-01	AW806881.1	EST_HUMAN	QV4-ST0023-160400-172-a01 ST0023 Homo sapiens cDNA
9602	22102	35065	4.6	5.1E-01	J05412.1	NT	Human regenerating protein (reg) gene, complete cds
9605	22105	35068	3.4	5.1E-01	W22302.1	EST_HUMAN	65B1 Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional
10065	22560	35555	0.95	5.1E-01	M94579.1	NT	Human carboxyl ester lipase (CEL) gene, complete cds
11874	24805		2.04	5.1E-01	BF030207.1	EST_HUMAN	601556863F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3826767 5'
12129	24385		2.01	5.1E-01	BF439982.1	EST_HUMAN	nac51f10.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3406218 3' similar to contains element TAR1 repetitive element ;
2180	14757	27326	1.4	5.0E-01	4885552	NT	Homo sapiens postmeiotic segregation increased 2-like 9 (PMS2L9), mRNA
2180	14757	27327	1.4	5.0E-01	4885552	NT	Homo sapiens postmeiotic segregation increased 2-like 9 (PMS2L9), mRNA
2189	14765	27335	5.46	5.0E-01	AF008210.1	NT	Buchnera aphidicola genomic fragment containing (chaperone Hsp60) groEL, DNA biosynthesis initiating protein (dnaA), ATP operon (atpCDGAHFEB), and putative chromosome replication protein (gidA) genes, complete cds; and termination factor Rho (rho) gene>
2189	14765	27336	5.46	5.0E-01	AF008210.1	NT	Buchnera aphidicola genomic fragment containing (chaperone Hsp60) groEL, DNA biosynthesis initiating protein (dnaA), ATP operon (atpCDGAHFEB), and putative chromosome replication protein (gidA) genes, complete cds; and termination factor Rho (rho) gene>
3740	16341	28809	5.58	5.0E-01	AE001785.1	NT	Thermotoga maritima section 97 of 136 of the complete genome
3811	16410	28875	0.65	5.0E-01	U55574.1	NT	Mus musculus anti-DNA immunoglobulin light chain IgM mRNA, antibody 363p.138, partial cds
3942	16540	29006	3.11	5.0E-01	AB033010.1	NT	Homo sapiens mRNA for KIAA1184 protein, partial cds
8467	21007		1.78	5.0E-01	M92304.1	NT	Xenopus laevis smooth muscle beta-tropomyosin mRNA, complete cds
8604	21143	34057	0.64	5.0E-01	BF107848.1	EST_HUMAN	601823850R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043485 3'
9379	20318	33219	3.1	5.0E-01	BF317212.1	EST_HUMAN	601903871F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4136632 5'
9543	22043	35004	1.34	5.0E-01	P35573	SWISSPROT	GLYCOGEN DEBRANCHING ENZYME (GLYCOGEN DEBRANCHER) [INCLUDES: 4-ALPHA-GLUCANOTRANSFERASE (OLIGO-1,4-1,4-GLUCANTRANSFERASE); AMYLO-1,6-GLUCOSIDASE (DEXTRIN 6-ALPHA-D-GLUCOSIDASE)]

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9543	22043	35005	1.34	5.0E-01	P35573	SWISSPROT	GLYCOGEN DEBRANCHING ENZYME (GLYCOGEN DEBRANCHER) [INCLUDES: 4-ALPHA-GLUCANOTRANSFERASE (OLIGO-1,4-1,4-GLUCANTRANSFERASE); AMYLO-1,6-GLUCOSIDASE (DEXTRIN 6-ALPHA-D-GLUCOSIDASE)]
10291	22786		1.04	5.0E-01	BE869218.1	EST_HUMAN	601445024F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849436 5'
11815	24187		3.45	5.0E-01	AF029215.1	NT	Mus musculus MRC OX-2 antigen homolog gene, exons 2-5, and complete cds
12554	24656		2.38	5.0E-01	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
12569	24668		4.27	5.0E-01	O13961	SWISSPROT	NUCLEAR ENVELOPE PROTEIN CUT11
822	13439	25946	2.31	4.9E-01	BF571462.1	EST_HUMAN	602076649F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4243860 5'
1699	14292	26827	1.6	4.9E-01	AJ243955.1	NT	Xenopus laevis mRNA for c-Jun protein, 1978 BP
1949	14533	27089	1.35	4.9E-01	U40869.1	NT	Cavia porcellus pulmonary surfactant protein A (SP-a) mRNA, complete cds
5602	18231	30681	1.32	4.9E-01	Q61554	SWISSPROT	FIBRILLIN 1 PRECURSOR
6187	18797	31565	2.35	4.9E-01	AF020931.1	NT	Homo sapiens diacylglycerol kinase 3 (DAGK3) gene, exon 10
6187	18797	31566	2.35	4.9E-01	AF020931.1	NT	Homo sapiens diacylglycerol kinase 3 (DAGK3) gene, exon 10
7475	19997	32862	1.9	4.9E-01	AB040051.1	NT	Oryza sativa subsp. japonica mEF-G mRNA for mitochondrial elongation factor G, complete cds
8920	21458		1.49	4.9E-01	BF209791.1	EST_HUMAN	601874964F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4102503 5'
9115	21651	34592	0.96	4.9E-01	AW339905.1	EST_HUMAN	hc90c02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2907266 3' similar to TR:O95714 O95714 HERC2. ;
9220	25125		2.2	4.9E-01	10946863	NT	Mus musculus unc13 homolog (C. elegans) 1 (Unc13h1), mRNA
10220	22715	35706	0.74	4.9E-01	AF053980.1	NT	Mus musculus adenylyl cyclase 1 (Adcy1) cDNA, partial cds
11704	24117		2.46	4.9E-01	AF176912.1	NT	Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene, complete cds
12546	25081		5.73	4.9E-01	AA613562.1	EST_HUMAN	nq22e11.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1144652 3'
12555	24657	30872	1.74	4.9E-01	AL163301.2	NT	Homo sapiens chromosome 21 segment HS21C101
12630	24708		1.36	4.9E-01	11431438	NT	Homo sapiens eukaryotic translation initiation factor 4 gamma, 1 (EIF4G1), mRNA
3591	16195		1.05	4.8E-01	AA912842.1	EST_HUMAN	ol32a09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1525144 3'
4782	17011		0.62	4.8E-01	4504850	NT	Homo sapiens potassium channel, subfamily K, member 5 (TASK-2) (KCNK5) mRNA, and translated products
5698	18324	30827	8.6	4.8E-01	J02987.1	NT	Saccharomyces cerevisiae) sporulation protein (SPO11) gene required for meiotic recombination, complete cds
6790	19381		4.22	4.8E-01	AA659878.1	EST_HUMAN	nu85f09.s1 NCI_CGAP_Alv1 Homo sapiens cDNA clone IMAGE:1217513
7357	19883		1.85	4.8E-01	5031650	NT	Homo sapiens reproduction 8 (D8S2298E) mRNA
7662	20174	33061	0.87	4.8E-01	AL163209.2	NT	Homo sapiens chromosome 21 segment HS21C009
7738	20246	33138	3.72	4.8E-01	AL161492.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 4
7738	20246	33139	3.72	4.8E-01	AL161492.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 4

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7846	20388	33291	1.36	4.8E-01	AI820744.1	EST_HUMAN	yj77f10.y5 Soares breast 2NbHBst Homo sapiens cDNA clone IMAGE:154795 5' similar to contains element MER6 repetitive element ;
9169	21746		1.13	4.8E-01	BE155148.1	EST_HUMAN	PM1-HT0350-201299-004-b04 HT0350 Homo sapiens cDNA
9921	22417		0.58	4.8E-01	BF568633.1	EST_HUMAN	602184267F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300048 5'
10607	23141		2.02	4.8E-01	X83502.1	NT	S.cerevisiae ORFs from chromosome X
11786	24170		1.29	4.8E-01	AL163227.2	NT	Homo sapiens chromosome 21 segment HS21C027
12016	24842		3.04	4.8E-01	AF227565.1	NT	Trypanosoma cruzi transposon VIP II SIRE repeat region
12646	24895		3.36	4.8E-01	AJ132984.1	NT	Chlamydomonas reinhardtii cop gene, exons 1-8
6638	19234	32036	8.72	4.7E-01	BF217173.1	EST_HUMAN	601883880F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4096387 5'
7107	19447	32263	0.78	4.7E-01	AI204374.1	EST_HUMAN	qf72a09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755544 3'
7806	20349	33257	0.52	4.7E-01	T11414.1	EST_HUMAN	hbc811 Human pancreatic islet Homo sapiens cDNA clone hbc811 5'end
7806	20349	33258	0.52	4.7E-01	T11414.1	EST_HUMAN	hbc811 Human pancreatic islet Homo sapiens cDNA clone hbc811 5'end
9005	21542	34473	0.5	4.7E-01	6981501	NT	Rattus norvegicus Spermine binding protein (Sbp), mRNA
10467	22961	35972	0.79	4.7E-01	AW087791.1	EST_HUMAN	xb69e11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2581580 3'
10727	23253		4.94	4.7E-01	AF102673.1	NT	Influenza A virus isolate hk51697 hemagglutinin (HA) gene, partial cds
10963	23478	36503	2.19	4.7E-01	U41069.1	NT	Human collagen alpha2(XI) (COL11A2) gene, exons 6 through 16, and partial cds
11163	23670	36715	11.61	4.7E-01	BF529658.1	EST_HUMAN	602043889F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4181303 5'
11254	23784	36840	2.89	4.7E-01	AW889448.1	EST_HUMAN	RC6-NT0029-240400-011-E08 NT0029 Homo sapiens cDNA
11904	24243		1.92	4.7E-01	BE887763.1	EST_HUMAN	601511333F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912488 5'
12036	24325		1.33	4.7E-01	AW341561.1	EST_HUMAN	hd11c08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2909198 3'
12666	24736		1.38	4.7E-01	AP000007.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 1485001-1738505 nt. position (7/7)
3797	16397	28862	2.23	4.6E-01	AW818638.1	EST_HUMAN	RC1-ST0278-040400-018-b06 ST0278 Homo sapiens cDNA
3806	16406	28870	1.68	4.6E-01	BF693300.1	EST_HUMAN	602081103F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4245481 5'
3806	16406	28871	1.68	4.6E-01	BF693300.1	EST_HUMAN	602081103F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4245481 5'
5323	17885		1.03	4.6E-01	M11267.1	NT	Bovine steroid 21-hydroxylase gene (P-450-c21) gene, complete cds
5429	17986	30390	22.08	4.6E-01	AL163248.2	NT	Homo sapiens chromosome 21 segment HS21C048
5440	17995	30400	1.37	4.6E-01	P51170	SWISSPROT	AMILORIDE-SENSITIVE SODIUM CHANNEL GAMMA-SUBUNIT (EPITHELIAL NA+ CHANNEL GAMMA SUBUNIT) (GAMMA ENAC) (NONVOLTAGE-GATED SODIUM CHANNEL 1 GAMMA SUBUNIT) (SCNEG) (GAMMA NACH)
5612	18241	30690	1.12	4.6E-01	BF313593.1	EST_HUMAN	601900234F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129472 5'
5612	18241	30691	1.12	4.6E-01	BF313593.1	EST_HUMAN	601900234F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129472 5'
5663	18290	30768	3.27	4.6E-01	Q90643	SWISSPROT	INTERFERON REGULATORY FACTOR 3 (IRF-3)
5663	18290	30769	3.27	4.6E-01	Q90643	SWISSPROT	INTERFERON REGULATORY FACTOR 3 (IRF-3)
5735	18361	31067	2.39	4.6E-01	BE734781.1	EST_HUMAN	601568755F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3843637 5'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5748	18374	31082	4.22	4.6E-01	AI247679.1	EST_HUMAN	qh59h02.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1849011 3' similar to TR:O15338 O15338 BUTYROPHILIN. ;
5748	18374	31083	4.22	4.6E-01	AI247679.1	EST_HUMAN	qh59h02.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1849011 3' similar to TR:O15338 O15338 BUTYROPHILIN. ;
5756	18382	31094	1.4	4.6E-01	P20050	SWISSPROT	MEIOSIS SPECIFIC PROTEIN HOP1
5828	18452		1.05	4.6E-01	AF212124.1	NT	Anolis schwartzi cytochrome b gene, partial cds; mitochondrial gene for mitochondrial product
5907	18529		0.86	4.6E-01	BE817247.1	EST_HUMAN	PM0-BN0260-120600-001-F07 BN0260 Homo sapiens cDNA
6058	18675	31417	0.75	4.6E-01	D26215.1	NT	Unidentified soil bacteria 16S rRNA gene encoding 16S ribosomal RNA
6404	19007	31788	1.05	4.6E-01	AE000894.1	NT	Methanobacterium thermoautotrophicum from bases 1165751 to 1176238 (section 100 of 148) of the complete genome
6865	19599	32429	1.36	4.6E-01	U62332.1	NT	Emericella nidulans NEMPA (nempA) gene, mitochondrial gene encoding putative mitochondrial protein, complete cds
6865	19599	32430	1.36	4.6E-01	U62332.1	NT	Emericella nidulans NEMPA (nempA) gene, mitochondrial gene encoding putative mitochondrial protein, complete cds
7712	20221	33108	0.86	4.6E-01	AA493577.1	EST_HUMAN	nh04h05.s1 NCI_CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943353 similar to contains Alu repetitive element;contains element L1 repetitive element ;
8262	20803	33721	13.23	4.6E-01	BF697399.1	EST_HUMAN	602130953F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287828 5'
9225	21741	34684	1.04	4.6E-01	P55202	SWISSPROT	ATRIAL NATRIURETIC PEPTIDE RECEPTOR B PRECURSOR (ANP-B) (ANPRB) (GC-B) (GUANYLATE CYCLASE)
9225	21741	34685	1.04	4.6E-01	P55202	SWISSPROT	ATRIAL NATRIURETIC PEPTIDE RECEPTOR B PRECURSOR (ANP-B) (ANPRB) (GC-B) (GUANYLATE CYCLASE)
9583	22083	35046	0.55	4.6E-01	AF162283.1	NT	Glycine max acetyl-CoA carboxylase (accB-1) gene, complete cds; nuclear gene for chloroplast product
9583	22083	35047	0.55	4.6E-01	AF162283.1	NT	Glycine max acetyl-CoA carboxylase (accB-1) gene, complete cds; nuclear gene for chloroplast product
9886	22383	35358	2.63	4.6E-01	AI915634.1	EST_HUMAN	wg73e12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2370766 3'
9886	22383	35359	2.63	4.6E-01	AI915634.1	EST_HUMAN	wg73e12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2370766 3'
10870	23391		3.09	4.6E-01	P98163	SWISSPROT	PUTATIVE VITELLOGENIN RECEPTOR PRECURSOR (YL)
10879	23400	36416	4.13	4.6E-01	BE185449.1	EST_HUMAN	IL5-HT0730-100500-075-g05 HT0730 Homo sapiens cDNA
10879	23400	36417	4.13	4.6E-01	BE185449.1	EST_HUMAN	IL5-HT0730-100500-075-g05 HT0730 Homo sapiens cDNA
11346	23044	36054	5.52	4.6E-01	AF019369.1	NT	Human thiopurine methyltransferase (TPMT) gene, exon 10 and complete cds
11346	23044	36055	5.52	4.6E-01	AF019369.1	NT	Human thiopurine methyltransferase (TPMT) gene, exon 10 and complete cds
12654	24726		1.26	4.6E-01	M22360.1	NT	Rat plasma proteinase inhibitor alpha-1-inhibitor III group 3 variants 6J, 12J, 13J, and 17J mRNA, partial cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1954	14538	27094	1.69	4.5E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
1954	14538	27095	1.69	4.5E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2695	15512	27982	4.77	4.5E-01	AA677086.1	EST_HUMAN	zj55d02.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:454179 3'
3346	15956	28431	0.64	4.5E-01	AW083761.1	EST_HUMAN	xc25c06.x1 NCI_CGAP_Co19 Homo sapiens cDNA clone IMAGE:2585290 3' similar to gb:L07807 DYNAMIN-1 (HUMAN);
3346	15956	28432	0.64	4.5E-01	AW083761.1	EST_HUMAN	xc25c06.x1 NCI_CGAP_Co19 Homo sapiens cDNA clone IMAGE:2585290 3' similar to gb:L07807 DYNAMIN-1 (HUMAN);
3359	15967	28444	5.18	4.5E-01	Q05793	SWISSPROT	BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC)
3424	16032	28512	1.15	4.5E-01	AF126378.1	NT	Mus musculus DNA polymerase epsilon catalytic subunit (Pole) gene, exons 2 through 12
4100	16694		1.35	4.5E-01	Q28247	SWISSPROT	COLLAGEN ALPHA 5(IV) CHAIN
4149	16741	29195	0.73	4.5E-01	AI708908.1	EST_HUMAN	as96e09.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone IMAGE:2353480 3'
4255	18015		4.04	4.5E-01	AW873495.1	EST_HUMAN	ho60g02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:3041810 3'
5078	17651	30092	1.16	4.5E-01	BE963445.2	EST_HUMAN	601657225R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866023 3'
5427	17984		26.74	4.5E-01	AF060195.1	NT	Mus musculus proteasome regulator PA28 beta subunit gene, complete cds
5737	18363	31070	1.37	4.5E-01	AW608814.1	EST_HUMAN	QV2-PT0012-140100-031-c09 PT0012 Homo sapiens cDNA
6719	19313		1.36	4.5E-01	Q00956	SWISSPROT	COAT PROTEIN
7443	19967	32834	1.69	4.5E-01	M37036.1	NT	Rat nucleolar proteins B23.1 and B23.2
7604	20117	32993	2.53	4.5E-01	AI858849.1	EST_HUMAN	wl32e02.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2426618 3' similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;
8249	20790		0.97	4.5E-01	M32661.1	NT	D.melanogaster Shaw2 protein mRNA, complete cds
8342	20883	33804	4.02	4.5E-01	AI648596.1	EST_HUMAN	tz56g11.x1 NCI_CGAP_Ov35 Homo sapiens cDNA clone IMAGE:2292644 3'
8494	21033	33954	0.69	4.5E-01	Q52728	SWISSPROT	POLY-BETA-HYDROXYBUTYRATE POLYMERASE (POLY(3-HYDROXYBUTYRATE) POLYMERASE) (PHB POLYMERASE) (PHB SYNTHASE) (POLY(3-HYDROXYALKANOATE) POLYMERASE) (PHA POLYMERASE) (PHA SYNTHASE) (POLYHYDROXYALKANOIC ACID SYNTHASE)
8716	21255		1.74	4.5E-01	11444786	NT	Homo sapiens hypothetical protein DKFZp547G183 (DKFZp547G183), mRNA
8929	21467	34385	0.69	4.5E-01	AE000218.1	NT	Escherichia coli K-12 MG1655 section 108 of 400 of the complete genome
9853	22351		1.02	4.5E-01	9630816	NT	Bombyx mori nuclear polyhedrosis virus, complete genome
10389	22883	35877	23.95	4.5E-01	M86006.1	EST_HUMAN	EST02531 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCY17
10389	22883	35878	23.95	4.5E-01	M86006.1	EST_HUMAN	EST02531 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCY17
10744	23268	36285	3.01	4.5E-01	AW591271.1	EST_HUMAN	xo14h01.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2703985 3' similar to SW:INT6_MOUSE Q64252 VIRAL INTEGRATION SITE PROTEIN INT-6. [1] ;
11131	23639		1.9	4.5E-01	AV719382.1	EST_HUMAN	AV719382 GLC Homo sapiens cDNA clone GLCCED12 5'

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11376	23828	36890	1.68	4.5E-01	BE066472.1	EST_HUMAN	RC3-BT0333-160300-016-e03 BT0333 Homo sapiens cDNA
11671	25070		3.3	4.5E-01	BE871461.1	EST_HUMAN	601449201F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852961 5'
12370	24540		2.13	4.5E-01	BF337531.1	EST_HUMAN	602035275F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183290 5'
12442	24578		6.25	4.5E-01	11422099	NT	Homo sapiens testis-specific kinase 2 (TESK2), mRNA
2081	14662		1.39	4.4E-01	6680503	NT	Mus musculus integral membrane-associated protein 1 (Itmap1), mRNA
2432	14999	27572	3.26	4.4E-01	P49765	SWISSPROT	VASCULAR ENDOTHELIAL GROWTH FACTOR B PRECURSOR (VEGF-B) (VEGF RELATED FACTOR)
3357	15965	28442	1.27	4.4E-01	AF058790.1	NT	Rattus norvegicus SynGAP-b mRNA, complete cds
3357	15965	28443	1.27	4.4E-01	AF058790.1	NT	Rattus norvegicus SynGAP-b mRNA, complete cds
3361	15969	28446	2.31	4.4E-01	BF056726.1	EST_HUMAN	7j91d02.y1 NCI_CGAP_Br16 Homo sapiens cDNA clone IMAGE:3393795 5'
4318	16904		1.28	4.4E-01	BE378707.1	EST_HUMAN	601237139F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3609393 5'
5134	17706		2.07	4.4E-01	BE141396.1	EST_HUMAN	MR0-HT0078-131299-007-g05 HT0078 Homo sapiens cDNA
5277	17839	30265	0.94	4.4E-01	U61154.1	NT	Buzura suppressaria nucleopolyhedrosisvirus ecdysteroid UDP-glucosyltransferase (egt) gene, complete cds
5417	17974		0.9	4.4E-01	AW814885.1	EST_HUMAN	MR1-ST0206-120400-022-g07 ST0206 Homo sapiens cDNA
5613	18242	30692	4.06	4.4E-01	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
5613	18242	30693	4.06	4.4E-01	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
5867	18489	31215	1.72	4.4E-01	S65019.1	NT	mucin [rats, Sprague-Dawley, sulfur-dioxide-treated tracheal epithelium, mRNA Partial, 390 nt]
5883	18505	31231	1.9	4.4E-01	AV720408.1	EST_HUMAN	AV720408 GLC Homo sapiens cDNA clone GLCCSC12 5'
6108	18724	31476	1.53	4.4E-01	AI198413.1	EST_HUMAN	qi62h11.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1861125 3' similar to TR:Q29168 Q29168 UNKNOWN PROTEIN ;
6108	18724	31477	1.53	4.4E-01	AI198413.1	EST_HUMAN	qi62h11.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1861125 3' similar to TR:Q29168 Q29168 UNKNOWN PROTEIN ;
6387	18990	31771	1.69	4.4E-01	AW080795.1	EST_HUMAN	xc27e08.x1 NCI_CGAP_Co18 Homo sapiens cDNA clone IMAGE:2585510 3' similar to TR:O95154 O95154 AFLATOXIN B1-ALDEHYDE REDUCTASE. ;
6470	19071		1.02	4.4E-01	AA776132.1	EST_HUMAN	ae85d11.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:970965 3' similar to gb:M16038 TYROSINE-PROTEIN KINASE LYN (HUMAN);
7429	19953	32818	0.89	4.4E-01	AE000571.1	NT	Helicobacter pylori 26695 section 49 of 134 of the complete genome
7782	20325		10.05	4.4E-01	Z11679.1	NT	S.tuberosum mRNA for induced stolon tip protein (partial)
8698	21237	34160	1.01	4.4E-01	AA056427.1	EST_HUMAN	zl69a03.s1 Stratagene colon (#937204) Homo sapiens cDNA clone IMAGE:509836 3'
9078	21614	34549	0.76	4.4E-01	AF112540.1	NT	HIV-1 isolate 08107v6 from USA, envelope glycoprotein (env) gene, partial cds
9111	21647	34587	0.56	4.4E-01	AW612578.1	EST_HUMAN	hh05c08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2954222 3' similar to SW:MSH6_HUMAN P52701 DNA MISMATCH REPAIR PROTEIN MSH6 ;
9214	21731	34674	1.13	4.4E-01	O62836	SWISSPROT	ZINC FINGER X-CHROMOSOMAL PROTEIN

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9872	22369	35347	1.69	4.4E-01	AI268650.1	EST_HUMAN	qo39f09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1910921 3'
9873	22370		2.12	4.4E-01	P28922	SWISSPROT	GLYCOPROTEIN B PRECURSOR (GLYCOPROTEIN 14)
10009	22504	35495	4.51	4.4E-01	P35590	SWISSPROT	TYROSINE-PROTEIN KINASE RECEPTOR TIE-1 PRECURSOR
10276	22771	35759	1.43	4.4E-01	S76404.1	NT	beta -HKA=H,K-ATPase beta-subunit [rats, Genomic, 8983 nt, segment 2 of 2]
10276	22771	35760	1.43	4.4E-01	S76404.1	NT	beta -HKA=H,K-ATPase beta-subunit [rats, Genomic, 8983 nt, segment 2 of 2]
11939	24271	31016	4.68	4.4E-01	6877874	NT	Mus musculus sodium channel, type X, alpha polypeptide (Scn10a), mRNA
11952	25000		14.98	4.4E-01	AL163282.2	NT	Homo sapiens chromosome 21 segment HS21C082
12517	24835		1.5	4.4E-01	P54725	SWISSPROT	UV EXCISION REPAIR PROTEIN PROTEIN RAD23 HOMOLOG A (HHR23A)
436	13069	25564	1.77	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
436	13069	25565	1.77	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
3096	15711	28182	0.91	4.3E-01	AW999477.1	EST_HUMAN	MR0-BN0070-270300-008-g04 BN0070 Homo sapiens cDNA
4231	16819	29268	1.21	4.3E-01	J00306.1	NT	Human somatostatin I gene and flanks
4495	13069	25564	3.96	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
4495	13069	25565	3.96	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
5567	18198	30645	0.76	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
5567	18198	30646	0.76	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
6049	18668	31407	1.34	4.3E-01	BE181655.1	EST_HUMAN	QV1-HT0638-070500-191-d08 HT0638 Homo sapiens cDNA
6065	18682	31424	2.06	4.3E-01	AF179825.1	NT	Saimiri sciureus olfactory receptor (SSC186) gene, partial cds
6809	19400	32215	4.28	4.3E-01	AJ001678.1	NT	Coturnix coturnix japonica ifnG gene
6949	19526		0.78	4.3E-01	O33367	SWISSPROT	DNA GYRASE SUBUNIT B
7456	19980		1.76	4.3E-01	BF348001.1	EST_HUMAN	602023134F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4158296 5'
8366	20906		2.66	4.3E-01	U97040.1	NT	Methanococcus voltae flagella-related protein C-I (flaC-flal) genes, complete cds
9179	21756	34702	0.7	4.3E-01	Y14604.1	NT	Erwinia amylovora rcsV gene
9642	22142	35109	2.63	4.3E-01	AW630048.1	EST_HUMAN	hh74e10.y1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2968554 5'
9642	22142	35110	2.63	4.3E-01	AW630048.1	EST_HUMAN	hh74e10.y1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2968554 5'
10128	22623	35614	0.57	4.3E-01	AW170559.1	EST_HUMAN	xn63e05.x1 Soares_NHCeC_cervical_tumor Homo sapiens cDNA clone IMAGE:2698400 3' similar to TR:O00189 O00189 MU-ADAPTIN-RELATED PROTEIN 2. ;
10811	19616	32451	2.52	4.3E-01	AF075629.1	NT	Equus caballus microsatellite LEX027
11588	24031	37101	1.54	4.3E-01	AI874332.1	EST_HUMAN	tz64d04.x1 NCI_CGAP_Ov35 Homo sapiens cDNA clone IMAGE:2293351 3'
11632	18198	30645	1.55	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
11632	18198	30646	1.55	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
12616	24696		2.81	4.3E-01	AJ003022.1	NT	Streptomyces coelicolor whiH gene
1402	15440	26524	1.39	4.2E-01	Q39102	SWISSPROT	CELL DIVISION PROTEIN FTSH HOMOLOG PRECURSOR
1991	14573		1.04	4.2E-01	AA761653.1	EST_HUMAN	nz24e09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1288696 3'

Table 4

Single Exon Probes Expressed In Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2068	14648		1.37	4.2E-01	AF258325.1	NT	Plasmodium falciparum multidrug resistance protein Pgh1 gene, complete cds
3669	16270	28738	4.91	4.2E-01	AE003947.1	NT	Xylella fastidiosa, section 93 of 229 of the complete genome
3699	16300	28768	1	4.2E-01	AI280338.1	EST_HUMAN	qi94b01.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1879945 3'
3773	18014		0.6	4.2E-01	N81203.1	EST_HUMAN	788iE1 fetal brain cDNA Homo sapiens cDNA clone 788iE1-K similar to R07879, Z40498
3948	16546	29014	0.73	4.2E-01	AW835527.1	EST_HUMAN	QV0-LT0015-180200-127-h01 LT0015 Homo sapiens cDNA
4054	16651	29118	0.98	4.2E-01	Q04886	SWISSPROT	SOX-8 PROTEIN
4807	17385	29835	4.3	4.2E-01	AA534093.1	EST_HUMAN	nj69h01.s1 NCI_CGAP_Pr10 Homo sapiens cDNA clone IMAGE:997777 similar to gb:M33600 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 BETA CHAIN (HUMAN);
4895	17470	29926	4.04	4.2E-01	R13467.1	EST_HUMAN	yf77e01.r1 Soares Infant brain 1NIB Homo sapiens cDNA clone IMAGE:28278 5'
5232	17786		3.77	4.2E-01	U50871.1	NT	Human familial Alzheimer's disease (STM2) gene, complete cds
5891	18514	31241	1.52	4.2E-01	BF242055.1	EST_HUMAN	601879721F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4108493 5'
5953	18575	31309	2.16	4.2E-01	AW854162.1	EST_HUMAN	RC3-CT0254-060400-029-g04 CT0254 Homo sapiens cDNA
6352	18957	31736	1.06	4.2E-01	AL163247.2	NT	Homo sapiens chromosome 21 segment HS21C047
7031	19565	32392	10.29	4.2E-01	AU158472.1	EST_HUMAN	AU158472 PLACE2 Homo sapiens cDNA clone PLACE2000470 3'
7031	19565	32393	10.29	4.2E-01	AU158472.1	EST_HUMAN	AU158472 PLACE2 Homo sapiens cDNA clone PLACE2000470 3'
7082	24776	32492	1.97	4.2E-01	S82504.1	NT	Brca1=breast cancer gene [rats, WF, spleen, Genomic, 419 nt, segment 2 of 2]
7150	19683	32524	5.81	4.2E-01	AL161547.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 47
7934	20476	33385	2.61	4.2E-01	AW957448.1	EST_HUMAN	EST369413 MAGE resequences, MAGE Homo sapiens cDNA
7934	20476	33386	2.61	4.2E-01	AW957448.1	EST_HUMAN	EST369413 MAGE resequences, MAGE Homo sapiens cDNA
8148	20689	33602	0.55	4.2E-01	4758039	NT	Homo sapiens cytochrome c oxidase subunit VIc (COX6C), nuclear gene encoding mitochondrial protein, mRNA
9235	21761	34706	0.52	4.2E-01	U57431.1	NT	Human cytomegalovirus early phosphoprotein p50 mRNA, complete cds
9235	21761	34707	0.52	4.2E-01	U57431.1	NT	Human cytomegalovirus early phosphoprotein p50 mRNA, complete cds
9880	22377		0.81	4.2E-01	AA705007.1	EST_HUMAN	zj95f01.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:462649 3'
10083	22578	35571	0.5	4.2E-01	AF181854.1	NT	Lassa virus strain 803213 glycoprotein precursor and nucleoprotein genes, complete cds
10390	22884	35879	1.35	4.2E-01	AW863686.1	EST_HUMAN	MR3-SN0010-280300-103-h07 SN0010 Homo sapiens cDNA
10921	23440	36461	3.69	4.2E-01	AB023489.1	NT	Oryzias latipes OIGC7 mRNA for membrane guanylyl cyclase, complete cds
11273	23726	36780	2.65	4.2E-01	BE966485.2	EST_HUMAN	601660352R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906085 3'
12561	24661		1.49	4.2E-01	AV731815.1	EST_HUMAN	AV731815 HTF Homo sapiens cDNA clone HTFBHH05 5'
1133	13736	26245	1.59	4.1E-01	AI905481.1	EST_HUMAN	RC-BT091-210199-142 BT091 Homo sapiens cDNA
1142	13745	26254	1.54	4.1E-01	AV705243.1	EST_HUMAN	AV705243 ADB Homo sapiens cDNA clone ADBAHF08 5'
1142	13745	26255	1.54	4.1E-01	AV705243.1	EST_HUMAN	AV705243 ADB Homo sapiens cDNA clone ADBAHF08 5'
2735	15290	27858	1.58	4.1E-01	7705283	NT	Homo sapiens anaphase-promoting complex subunit 7 (APC7), mRNA
2987	15582	28061	2.11	4.1E-01	AL161536.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 36

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2967	15582	28062	2.11	4.1E-01	AL161536.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 36
3342	15952	28428	0.66	4.1E-01	AA906344.1	EST_HUMAN	oj94b08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1505943 3'
3839	16438	28899	0.58	4.1E-01	AW961292.1	EST_HUMAN	EST373364 MAGE resequences, MAGG Homo sapiens cDNA
3839	16438	28900	0.58	4.1E-01	AW961292.1	EST_HUMAN	EST373364 MAGE resequences, MAGG Homo sapiens cDNA
4361	16948	29390	2.82	4.1E-01	AJ249207.1	NT	Rhodococcus sp. AD45 isoG, isoH, isoI, isoJ, isoA, isoB, isoC, IsoD, IsoE and isoF genes
4393	16979		0.76	4.1E-01	AA909257.1	EST_HUMAN	om33d02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1542819 3'
4774	17355	29807	1.31	4.1E-01	AV747880.1	EST_HUMAN	AV747880 NPC Homo sapiens cDNA clone NPCBDF10 5'
6141	18755	31513	3.97	4.1E-01	BF681393.1	EST_HUMAN	602156590F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4297319 5'
7460	19983	32848	2.74	4.1E-01	U67535.1	NT	Methanococcus jannaschii section 77 of 150 of the complete genome
7979	20521	33427	1.31	4.1E-01	BF574604.1	EST_HUMAN	602133261F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288238 5'
9019	21556	34484	1.26	4.1E-01	6755521	NT	Mus musculus signaling intermediate in Toll pathway-evolutionarily conserved (Sitpec-pending), mRNA
9484	21941		0.61	4.1E-01	AF160597.1	NT	Voelavo gymnocaudus Vgym560 cytochrome b (cytb) gene, complete cds; mitochondrial gene for mitochondrial product
10164	22659		1.26	4.1E-01	AL139076.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 3/6
10310	22804	35796	0.79	4.1E-01	AV649579.1	EST_HUMAN	AV649579 GLC Homo sapiens cDNA clone GLCBVD12 3'
10401	22895	35890	0.51	4.1E-01	P18584	SWISSPROT	PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (59 KDA IMMUNOGENIC PROTEIN) (SK59)
10401	22895	35891	0.51	4.1E-01	P18584	SWISSPROT	PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (59 KDA IMMUNOGENIC PROTEIN) (SK59)
10471	22965		2.29	4.1E-01	BF349382.1	EST_HUMAN	CM2-HT0137-200999-010-e08 HT0137 Homo sapiens cDNA
10719	23247	36262	45.22	4.1E-01	X58700.1	NT	Zea mays ZMPMS2 gene for 19 kDa zein protein
11270	23008	36015	3.57	4.1E-01	Q09470	SWISSPROT	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV1.1 (HUKI) (HBK1)
12290	25049		2.6	4.1E-01	D87675.1	NT	Homo sapiens DNA for amyloid precursor protein, complete cds
147	15408		4.55	4.0E-01	AW847123.1	EST_HUMAN	RC2-CT0201-290999-012-d10 CT0201 Homo sapiens cDNA
1077	13681	26191	0.82	4.0E-01	8404656	NT	Laqueus rubellus mitochondrion, complete genome
1384	13978	26505	1.51	4.0E-01	AF203478.1	NT	Drosophila melanogaster Dalmatian (dmt) mRNA, complete cds
1532	14124		4.1	4.0E-01	6679258	NT	Mus musculus platelet derived growth factor receptor, beta polypeptide (Pdgfrb), mRNA
2049	15456	27200	1.22	4.0E-01	Z96933.1	NT	Ascobolus immersus masc2 gene
2049	15456	27201	1.22	4.0E-01	Z96933.1	NT	Ascobolus immersus masc2 gene
2204	14780	27352	17.82	4.0E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2204	14780	27353	17.82	4.0E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2831	12811	25299	1.45	4.0E-01	6678490	NT	Mus musculus ubiquitin-protein ligase e3 componen n-recognin (Ubr1), mRNA
2995	15611	28090	1.23	4.0E-01	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
2995	15611	28091	1.23	4.0E-01	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3758	16359	28829	2.17	4.0E-01	AF068903.1	NT	Streptococcus pneumoniae YllC (yllC), YllD (yllD), penicillin-binding protein 2x (pbp2x), and undecaprenyl-phosphate-UDP-MurNAc-pentapeptide phospho-MurNAc-pentapeptide transferase (mraY) genes, complete cds
3899	16498	28960	3.04	4.0E-01	AJ277511.1	NT	Ovis aries partial JD2 gene for T cell receptor delta chain (TCRDJ2), exon 1
3899	16498	28961	3.04	4.0E-01	AJ277511.1	NT	Ovis aries partial JD2 gene for T cell receptor delta chain (TCRDJ2), exon 1
4942	17517		8.41	4.0E-01	Q31849	SWISSPROT	NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5, CHLOROPLAST
6069	18686	31429	1.16	4.0E-01	AW970610.1	EST_HUMAN	EST382691 MAGE resequences, MAGK Homo sapiens cDNA
6567	19165	31961	0.87	4.0E-01	P27285	SWISSPROT	STRUCTURAL POLYPROTEIN (P130) (CONTAINS: COAT PROTEIN C ; SPIKE GLYCOPROTEINS E3, E2 AND E1; 6 KD PEPTIDE)
7777	20286	33183	0.72	4.0E-01	P27548	SWISSPROT	MICROTUBULE-ASSOCIATED PROTEIN 4
7869	20411	33317	0.46	4.0E-01	BF092634.1	EST_HUMAN	MR4-TN0110-180900-202-g02 TN0110 Homo sapiens cDNA
7954	20496	33406	0.99	4.0E-01	AB016625.1	NT	Homo sapiens OCTN2 gene, complete cds
8938	21474	34394	0.98	4.0E-01	AA323289.1	EST_HUMAN	EST26066 Cerebellum II Homo sapiens cDNA 5' end similar to EST containing Alu repeat
11443	23893		1.65	4.0E-01	BF030262.1	EST_HUMAN	601558283F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3828092 5'
11568	24015		3.52	4.0E-01	L76080.1	NT	Synechocystis sp. PCC 9413 transposase gene, complete cds
11958	24901		2.5	4.0E-01	AL163300.2	NT	Homo sapiens chromosome 21 segment HS21C100
12518	24636		1.42	4.0E-01	P36049	SWISSPROT	HYPOTHETICAL 49.7 KD PROTEIN IN GIN2-STE3 INTERGENIC REGION
1420	14013	26543	1.98	3.9E-01	AF206618.1	NT	Gorilla gorilla carboxyl-ester lipase (CEL) gene, complete cds
2668	15226	27798	3.8	3.9E-01	AB033019.1	NT	Homo sapiens mRNA for KIAA1193 protein, partial cds
2730	15285	27851	3.79	3.9E-01	X82032.1	NT	H.sapiens B-myb gene
2730	15285	27852	3.79	3.9E-01	X82032.1	NT	H.sapiens B-myb gene
3131	15745	28214	3.95	3.9E-01	AJ225896.1	NT	Sinorhizobium meliloti egl, syrB2, cya3 genes and orf3
4153	16745	29199	1.49	3.9E-01	BF592611.1	EST_HUMAN	7i61d01.x1 NCI_CGAP_Br16 Homo sapiens cDNA clone IMAGE:3339169 3'
5130	17702	30136	1.86	3.9E-01	BE728667.1	EST_HUMAN	601563948F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3833699 5'
6090	18706	31454	6.44	3.9E-01	BF208036.1	EST_HUMAN	601862362F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082055 5'
6426	19029	31812	0.68	3.9E-01	U82695.2	NT	Homo sapiens zinc finger protein 92 (ZFP92), expressed-Xq28STS protein (XQ28ORF), and biglycan (BGN) genes, complete cds; and plasma membrane calcium ATPase isoform 3 (PMCA3) gene, partial cds
7896	20438	33343	0.78	3.9E-01	U79415.1	NT	Homo sapiens prepro dipeptidyl peptidase I (DPP-I) gene, complete cds
8795	21334	34259	0.73	3.9E-01	AW177011.1	EST_HUMAN	CM3-CT0105-170899-004-b08 CT0105 Homo sapiens cDNA
8804	21343		0.7	3.9E-01	BF348634.1	EST_HUMAN	602019944F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4155322 5'
9161	21696	34640	1.24	3.9E-01	AW195888.1	EST_HUMAN	xn86d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2701351 3' similar to TR:O94821 O94821 KIAA0713 PROTEIN ;

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9464	21989	34945	1.42	3.9E-01	AI937337.1	EST_HUMAN	wp76a02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2467658 3' similar to SW:RFX5_HUMAN P48382 BINDING REGULATORY FACTOR. ;
9792	22290	35274	3.68	3.9E-01	M19879.1	NT	Human clabindin 27 gene, exons 10 and 11, and L1 and Alu repeats
9856	22354		0.5	3.9E-01	11465620	NT	Porphyra purpurea mitochondrion, complete genome
10071	22566	35561	0.69	3.9E-01	D86722.1	NT	Nicotiana tabacum mRNA for TATA binding protein (TBP), complete cds
10492	22986	35993	0.46	3.9E-01	M18440.1	NT	Human beta-B2-crystallin (B2-1) gene, exon 4, partial cds
10700	23229		1.82	3.9E-01	AV695974.1	EST_HUMAN	AV695974 GKC Homo sapiens cDNA clone GKCBQC11 5'
11729	24977		3.42	3.9E-01	AF304354.1	NT	Homo sapiens proteoglycan 3 (PRG3) gene, complete cds
11854	24214		1.42	3.9E-01	Q61670	SWISSPROT	HOMEOBOX PROTEIN HLX1
11930	24265	31015	1.56	3.9E-01	AE001811.1	NT	Thermotoga maritima section 123 of 136 of the complete genome
12389	24551		1.37	3.9E-01	11433335	NT	Homo sapiens hypothetical protein FLJ10583 (FLJ10583), mRNA
171	12834		19.28	3.8E-01	7019488	NT	Homo sapiens protein kinase PKNbeta (pknbeta), mRNA
531	13162		3.11	3.8E-01	AB029291.1	NT	Mus musculus pcm-1 mRNA for pericentriolar material-1, complete cds
1911	14496		0.99	3.8E-01	AE003870.1	NT	Xylella fastidiosa, section 16 of 229 of the complete genome
2605	15167	27734	1.89	3.8E-01	AF214117.1	NT	Arabidopsis thaliana putative c-myb-like transcription factor (MYB3R-3) mRNA, complete cds
2661	15473	27791	3.94	3.8E-01	6678002	NT	Mus musculus solute carrier family 1, member 6 (Slc1a6), mRNA
3034	15650		0.89	3.8E-01	AJ251057.1	NT	Human immunodeficiency virus type 1 complete genome (isolate 98SE-MP1213)
3084	15699	28173	2.2	3.8E-01	AF043383.1	NT	Pleuronectes americanus aminopeptidase N (ampN) gene, partial cds
3530	16135	28615	9.83	3.8E-01	AL161518.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 30
3592	16196		0.59	3.8E-01	AI807219.1	EST_HUMAN	wf38b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357855 3'
3609	16196		0.75	3.8E-01	AI807219.1	EST_HUMAN	wf38b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357855 3'
3820	16420	28882	0.94	3.8E-01	BE154080.1	EST_HUMAN	PM0-HT0339-200400-010-G01 HT0339 Homo sapiens cDNA
3989	16587	29058	0.8	3.8E-01	6754095	NT	Mus musculus general transcription factor II I (Gtf2i), mRNA
4138	16730	29183	0.69	3.8E-01	AJ271361.2	NT	Takifugu rubripes wnt2 (partial), frank1, cftr and frank2 (partial) genes
5271	17833	30259	0.99	3.8E-01	BE544653.1	EST_HUMAN	601074110F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3460154 5'
5412	17969	30378	1.07	3.8E-01	H64927.1	EST_HUMAN	yr68a11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:210428 5' similar to gb|M87933|HUMAALU364 Human carcinoma cell-derived Alu RNA transcript, (rRNA); gb:M96956 EPIDERMAL GROWTH FACTOR-LIKE CRIPTO PROTEIN (HUMAN);contains Alu repetitive element;contains MER4 repetitive element ;
5794	18419	31135	1.11	3.8E-01	Q04888	SWISSPROT	TRANSCRIPTION FACTOR SOX-10
6481	19082		0.68	3.8E-01	S46825.1	NT	prion protein [mink, Genomic, 2446 nt]
6737	19331	32137	5.29	3.8E-01	BE072399.1	EST_HUMAN	QV3-BT0537-271299-049-e02 BT0537 Homo sapiens cDNA
6857	19591	32423	3.97	3.8E-01	AI374601.1	EST_HUMAN	ta54f11.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:2047917 3' similar to contains Alu repetitive element;

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7019	19517	32339	1.33	3.8E-01	AL161513.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 25
7525	20045		4.75	3.8E-01	X61597.1	NT	M.musculus gene for kallikrein-binding protein
8238	20779	33700	0.49	3.8E-01	M81385.1	NT	Mouse liver receptor homologous protein (LRH-1) mRNA, complete cds
8492	21031	33951	2.34	3.8E-01	AB046851.1	NT	Homo sapiens mRNA for KIAA1631 protein, partial cds
8560	21099	34019	1.14	3.8E-01	11441264	NT	Homo sapiens FOS-like antigen-1 (FOSL1), mRNA
8751	21290	34210	1.12	3.8E-01	AL163279.2	NT	Homo sapiens chromosome 21 segment HS21C079
9480	21879		4.03	3.8E-01	T95413.1	EST_HUMAN	ye43h06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:120539 5' similar to contains Alu repetitive element;contains PTR5 repetitive element ;
11408	23859		3.5	3.8E-01	BE719219.1	EST_HUMAN	RC0-HT0841-040800-032-b12 HT0841 Homo sapiens cDNA
11541	23989	37060	2.95	3.8E-01	R42550.1	EST_HUMAN	yf92h11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:30289 3'
11541	23989	37061	2.95	3.8E-01	R42550.1	EST_HUMAN	yf92h11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:30289 3'
11940	24272		2.81	3.8E-01	AE001124.1	NT	Borrelia burgdorferi (section 10 of 70) of the complete genome
12069	24998		1.75	3.8E-01	U94788.1	NT	Human p53 (TP53) gene, complete cds
12189	24421		1.45	3.8E-01	BE829256.1	EST_HUMAN	QV3-ET0063-190700-271-a05 ET0063 Homo sapiens cDNA
12565	24664		2.22	3.8E-01	U78031.1	NT	Mus musculus apoptosis inhibitor bcl-x (bcl-x) gene, exon 3 and complete cds
12638	24713	30866	1.25	3.8E-01	AF194972.1	NT	Mus musculus developmental control protein mRNA, partial cds
2521	15085	27657	15.01	3.7E-01	AB037831.1	NT	Homo sapiens mRNA for KIAA1410 protein, partial cds
3507	16112	28589	10.94	3.7E-01	AF056336.1	NT	Danio rerio bone morphogenetic protein 4 precursor (BMP4) gene, complete cds
3938	16536	29003	0.68	3.7E-01	AA319482.1	EST_HUMAN	EST21715 Adrenal gland tumor Homo sapiens cDNA 5' end
4313	16899	29344	9.19	3.7E-01	AI218707.1	EST_HUMAN	ok39c07.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:1510188 3'
4412	16997	29440	1.18	3.7E-01	AW876037.1	EST_HUMAN	MR3-OT0007-080300-104-b02 OT0007 Homo sapiens cDNA
4479	17064	29514	3.13	3.7E-01	AE002408.1	NT	Neisseria meningitidis serogroup B strain MC58 section 50 of 206 of the complete genome
5936	18557	31285	1.27	3.7E-01	AF135187.1	NT	Homo sapiens interferon-induced protein p78 (MX1) gene, complete cds
6105	18721	31474	0.94	3.7E-01	AL163278.2	NT	Homo sapiens chromosome 21 segment HS21C078
6633	19229	32033	1	3.7E-01	M10606.1	NT	Chicken (White leghorn) delta-1 and delta-2 crystallin genes, complete cds
6651	19247		0.81	3.7E-01	L10353.1	NT	Mus saxicola haptoglobin mRNA, complete cds
7197	19728	32579	4.44	3.7E-01	11525843	NT	Homo sapiens tumor endothelial marker 7 precursor (TEM7), mRNA
8271	20812	33733	1.88	3.7E-01	11436739	NT	Homo sapiens chromosome 12 open reading frame 4 (C12ORF4), mRNA
8271	20812	33734	1.88	3.7E-01	11436739	NT	Homo sapiens chromosome 12 open reading frame 4 (C12ORF4), mRNA
8308	20847	33770	0.76	3.7E-01	AA902912.1	EST_HUMAN	ok43b11.s1 NCI_CGAP_Lei2 Homo sapiens cDNA clone IMAGE:1516701 3'
9129	21664		1.54	3.7E-01	AJ271386.1	NT	Gallus gallus mRNA for beta-carotene 15,15'-dioxygenase (bCDO gene)
10074	22569		0.46	3.7E-01	K00691.1	NT	mouse ig germline alpha membrane exons region
10111	22606	35596	4.17	3.7E-01	AI336411.1	EST_HUMAN	qt46b07.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:1950997 3'
10909	23428	36446	3.47	3.7E-01	AJ297357.1	NT	Homo sapiens partial LIMD1 gene for LIM domains containing protein 1 and KIAA0851 gene

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10909	23428	36447	3.47	3.7E-01	AJ297357.1	NT	Homo sapiens partial LIMD1 gene for LIM domains containing protein 1 and KIAA0851 gene
11341	23039	36048	4.81	3.7E-01	X04122.1	NT	Bovine mRNA for terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31)
11549	23997		1.6	3.7E-01	AA973540.1	EST_HUMAN	oo46d03.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1569221 3' similar to gb:M77698 TRANSCRIPTIONAL REPRESSOR PROTEIN YY1 (HUMAN);
11603	24046		2.76	3.7E-01	6677678	NT	Mus musculus retinoblastoma 1 (Rb1), mRNA
11640	24601		1.82	3.7E-01	J04982.1	NT	Human heart/skeletal muscle ATP/ADP translocator (ANT1) gene, complete cds
11821	24191		4.15	3.7E-01	AJ243525.1	NT	Chlamydophila psittaci partial omp1 gene for outer membrane protein 1
11913	24251		4.72	3.7E-01	D86976.1	NT	Human mRNA for KIAA0223 gene, partial cds
12302	24499		2.94	3.7E-01	AL121154.1	EST_HUMAN	DKFZp762K075_r1 762 (synonym: hmel2) Homo sapiens cDNA clone DKFZp762K075 5'
12377	24545	30904	7.01	3.7E-01	Y18000.1	NT	Homo sapiens NF2 gene
282	12938	25424	1.07	3.6E-01	AJ009609.1	NT	Brassica napus mRNA for MAP4K alpha2 protein
1033	13643		8.45	3.6E-01	U89241.1	NT	Human mibp gene, partial cds
1357	13951	26477	4.32	3.6E-01	T80255.1	EST_HUMAN	yd03e05.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:24443 5'
1357	13951	26478	4.32	3.6E-01	T80255.1	EST_HUMAN	yd03e05.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:24443 5'
1959	14543	27099	6.39	3.6E-01	AW590184.1	EST_HUMAN	hg33f02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2947419 3'
1959	14543	27100	6.39	3.6E-01	AW590184.1	EST_HUMAN	hg33f02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2947419 3'
1994	14576	27136	7.23	3.6E-01	AF216207.1	NT	Mus musculus ribosomal protein S19 (Rps19) gene, complete cds
2098	14677		0.88	3.6E-01	AF056927.1	NT	Rattus norvegicus repeat element associated with the Rasgrf1 gene
2309	14881		1.13	3.6E-01	AB002321.1	NT	Human mRNA for KIAA0323 gene, partial cds
2433	15000		2.49	3.6E-01	X76725.1	NT	P.irregulare (P3804) gene for actin
2528	15092	27665	3.34	3.6E-01	AW812033.1	EST_HUMAN	RC5-ST0171-181099-011-g07 ST0171 Homo sapiens cDNA
2659	15218	27789	1.38	3.6E-01	P24206	SWISSPROT	PROTEIN-L-ISOASPARTATE O-METHYLTRANSFERASE (PROTEIN-BETA-ASPARTATE METHYLTRANSFERASE) (PIMT) (PROTEIN L-ISOASPARTYL METHYLTRANSFERASE) (L-ISOASPARTYL PROTEIN CARBOXYL METHYLTRANSFERASE)
2924	16012		10.38	3.6E-01	AF199485.1	NT	Drosophila melanogaster sugar transporter 3 (sut3) mRNA, complete cds
3516	16121	28600	2.16	3.6E-01	X76758.1	NT	H.sapiens serotonin transporter gene, exons 9 and 10
3516	16121	28601	2.16	3.6E-01	X76758.1	NT	H.sapiens serotonin transporter gene, exons 9 and 10
4497	17081	29530	1.97	3.6E-01	BE707883.1	EST_HUMAN	RC1-HT0545-150600-014-b12 HT0545 Homo sapiens cDNA
4850	17428	29880	0.94	3.6E-01	AJ009609.1	NT	Brassica napus mRNA for MAP4K alpha2 protein
4867	17443	29894	0.65	3.6E-01	Y11526.1	NT	Z.mays mRNA for casein kinase II alpha subunit
5153	17723	30154	2.28	3.6E-01	AW339393.1	EST_HUMAN	ha02g04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2872568 3'
5261	17824	30249	0.58	3.6E-01	BE067699.1	EST_HUMAN	MR4-BT0358-270300-005-c10 BT0358 Homo sapiens cDNA
6236	18845	31616	1.16	3.6E-01	P16431	SWISSPROT	FORMATE HYDROGENLYASE SUBUNIT 5 PRECURSOR (FHL SUBUNIT 5) (HYDROGENASE-3 COMPONENT E)

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6604	19201	32006	1.68	3.6E-01	Y10196.1	NT	Homo sapiens PHEX gene
7202	19733		4.57	3.6E-01	R94090.1	EST_HUMAN	yt74a06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:275967 5'
7327	19854	32717	1.73	3.6E-01	AW027174.1	EST_HUMAN	wt72c10.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone IMAGE:2513010 3' similar to TR:O15117 O15117 FYN BINDING PROTEIN. [1] ;
8166	20707	33623	0.68	3.6E-01	P98167	SWISSPROT	SCO-SPONDIN
8221	20762	33678	13.59	3.6E-01	AL161583.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 79
8932	21470	34388	3.06	3.6E-01	4504956	NT	Homo sapiens lysosomal-associated membrane protein 2 (LAMP2), transcript variant LAMP2A, mRNA
8932	21470	34389	3.06	3.6E-01	4504956	NT	Homo sapiens lysosomal-associated membrane protein 2 (LAMP2), transcript variant LAMP2A, mRNA
9120	21656	34597	1.32	3.6E-01	AL163204.2	NT	Homo sapiens chromosome 21 segment HS21C004
9320	21834	34784	0.92	3.6E-01	X17550.1	NT	D. melanogaster singed gene, exons 3, 4, 5 & 6
9320	21834	34785	0.92	3.6E-01	X17550.1	NT	D. melanogaster singed gene, exons 3, 4, 5 & 6
9390	21913		0.54	3.6E-01	X62825.1	NT	C.perfringens plc gene for phospholipase C upstream region containing bent DNA fragment
9777	22275	35260	16.15	3.6E-01	Q53194	SWISSPROT	PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TS
9904	22401	35374	0.53	3.6E-01	AW752901.1	EST_HUMAN	MR2-CT0222-211099-002-b10 CT0222 Homo sapiens cDNA
9904	22401	35375	0.53	3.6E-01	AW752901.1	EST_HUMAN	MR2-CT0222-211099-002-b10 CT0222 Homo sapiens cDNA
10823	23344	36359	2.51	3.6E-01	BE902390.1	EST_HUMAN	601676418F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958997 5'
10987	23501	36531	4.15	3.6E-01	AB004293.1	NT	Arabidopsis thaliana mRNA for SigB, complete cds
11065	23577	36615	2.02	3.6E-01	L41687.1	NT	Mus musculus T-cell receptor V region delta 1 chain gene, 5' region
11318	23016	36025	4.07	3.6E-01	AE000856.1	NT	Methanobacterium thermoautotrophicum from bases 702375 to 714311 (section 62 of 148) of the complete genome
11680	25109		2.45	3.6E-01	Y19210.1	NT	Homo sapiens hHb5 gene for hair keratin, exons 1 to 9
11768	24159		5.79	3.6E-01	AE000335.1	NT	Escherichia coli K-12 MG1655 section 225 of 400 of the complete genome
11923	24258		4.7	3.6E-01	U66888.1	NT	Mus musculus Emr1 mRNA, complete cds
12308	24502		2.16	3.6E-01	11432598	NT	Homo sapiens myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophila) homolog); translocated to, 10 (AF10), mRNA
120	12791	25273	1.35	3.5E-01	AL161536.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 36
223	12884	25369	2.67	3.5E-01	6678933	NT	Mus musculus mannose receptor, C type 2 (Mrc2), mRNA
708	13327	25814	4.48	3.5E-01	AL161581.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 77
754	13373	25868	1.39	3.5E-01	7706136	NT	Homo sapiens GAP-like protein (LOC51306), mRNA
754	13373	25869	1.39	3.5E-01	7706136	NT	Homo sapiens GAP-like protein (LOC51306), mRNA
812	13430	25935	3.83	3.5E-01	BF129796.1	EST_HUMAN	601811060R1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4053951 3'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1682	14274	26807	1.91	3.5E-01	U35776.1	NT	Rattus norvegicus ADP-ribosylation factor-directed GTPase activating protein mRNA, complete cds
2322	14893	27469	1.12	3.5E-01	P06798	SWISSPROT	HOMEOBOX PROTEIN HOX-A4 (HOX-1.4) (MH-3)
2637	15472	27770	1.92	3.5E-01	AA223252.1	EST_HUMAN	zr08a09.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone IMAGE:650872 3'
2729	15284		11.8	3.5E-01	U05897.1	NT	Fibrobacter succinogenes S85 endoglucanase E (celE) and endoglucanase D (celD) gene, complete cds
3030	15646	28124	0.57	3.5E-01	AA057691.1	EST_HUMAN	zl94f03.r1 Stratagene corneal stroma (#937222) Homo sapiens cDNA clone IMAGE:512285 5'
3878	16476		1.27	3.5E-01	AA642138.1	EST_HUMAN	nr60d03.s1 NCI_CGAP_Lym3 Homo sapiens cDNA clone IMAGE:1172357 3'
4349	16936	29377	2.3	3.5E-01	AF071253.1	NT	Danio rerio homeobox protein (hoxb5b) gene, complete cds
5014	17588	30031	0.57	3.5E-01	N81203.1	EST_HUMAN	788IE1 fetal brain cDNA Homo sapiens cDNA clone 788IE1-K similar to R07879, Z40498
5066	17639	30082	4.33	3.5E-01	M18349.1	NT	Rat leukocyte common antigen (L-CA) gene, exons 1 through 5
5537	18169	30583	0.74	3.5E-01	Q96687	SWISSPROT	EARLY E2A DNA-BINDING PROTEIN
5537	18169	30584	0.74	3.5E-01	Q96687	SWISSPROT	EARLY E2A DNA-BINDING PROTEIN
5738	18364	31071	1.42	3.5E-01	D42045.1	NT	Human mRNA for KIAA0086 gene, complete cds
6384	18988		0.9	3.5E-01	AW863916.1	EST_HUMAN	PM4-SN0012-030400-001-a11 SN0012 Homo sapiens cDNA
6540	19139	31932	0.75	3.5E-01	AA431833.1	EST_HUMAN	zw79f03.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782429 5' similar to TR:G1066935 G1066935 F10F2.1 ;
6578	19176	31976	0.72	3.5E-01	U37150.1	NT	Bos taurus peptide methionine sulfoxide reductase (msrA) mRNA, complete cds
6770	19363	32172	0.93	3.5E-01	O24357	SWISSPROT	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE, CHLOROPLAST PRECURSOR (G6PD)
7116	19456		3.51	3.5E-01	X98505.1	NT	S.scrofa mRNA for CD31 protein (PECAM-1)
8016	20558		2.02	3.5E-01	11448042	NT	Homo sapiens tumor protein p53-binding protein, 2 (TP53BP2), mRNA
8019	20561	33462	0.65	3.5E-01	BF358871.1	EST_HUMAN	RC4-ET0024-260600-014-d07 ET0024 Homo sapiens cDNA
8403	20943		0.61	3.5E-01	AF051561.1	NT	Rattus norvegicus Na-K-Cl cotransporter (Nkcc1) mRNA, complete cds
8857	21396	34319	1.12	3.5E-01	4507610	NT	Homo sapiens tyrosine kinase non-receceptor 1 (TNK1), mRNA
9651	22150	35120	1.93	3.5E-01	Q02294	SWISSPROT	VOLTAGE-DEPENDENT N-TYPE CALCIUM CHANNEL ALPHA-1B SUBUNIT (CALCIUM CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE ISOFORM 5) (BRAIN CALCIUM CHANNEL III) (BIII)
9801	22299	35284	4.91	3.5E-01	Z26825.1	NT	X.laevis gene for albumin including HP1 enhancer
9877	22374	35351	1.14	3.5E-01	BE174794.1	EST_HUMAN	QV2-HT0577-090400-128-c07 HT0577 Homo sapiens cDNA
10613	23148	36157	4	3.5E-01	X61084.1	NT	C.griseus rhodopsin gene for opsin protein
10902	23422	36440	2.09	3.5E-01	AJ243178.1	NT	Gallus gallus SPARC gene for osteonectin, promoter and exon 1
10902	23422	36441	2.09	3.5E-01	AJ243178.1	NT	Gallus gallus SPARC gene for osteonectin, promoter and exon 1
11462	23912	36979	1.93	3.5E-01	N77597.1	EST_HUMAN	yz90h12.r1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:290375 5'
11532	23980	37050	1.71	3.5E-01	L05145.1	NT	Human glucokinase (GCK) gene, repeat polymorphism

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11778	25112		1.51	3.5E-01	AF297468.1	NT	Schistosoma mansoni strain NMRI chromatin assembly factor 1 small subunit-like protein (RBAP48) mRNA, complete cds
11851	24211		7.56	3.5E-01	X64565.1	NT	B.taurus atpA1 gene for F(0)F(1) ATP synthase alpha-subunit
12014	24313		2.03	3.5E-01	AE001774.1	NT	Thermotoga maritima section 86 of 136 of the complete genome
12209	24433		2.21	3.5E-01	AE001691.1	NT	Thermotoga maritima section 3 of 136 of the complete genome
12643	24950	30625	2.64	3.5E-01	H80814.1	EST_HUMAN	ys64f11.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:219597 5'
12643	24950	30626	2.64	3.5E-01	H80814.1	EST_HUMAN	ys64f11.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:219597 5'
736	13356		1.5	3.4E-01	AJ242956.1	NT	Homo sapiens partial N-myc (exon 3), HPV45 L2, HPV45 L1, HPV45 E6, HPV45 E7 and HPV45 E1 genes isolated from IC4 cervical carcinoma cell line
1011	13621	26136	7.62	3.4E-01	Y09798.2	NT	Pseudomonas fluorescens colR, colS genes, orf222 and partial inaA gene
1013	13623	26138	8.97	3.4E-01	AW380120.1	EST_HUMAN	QV3-HT0261-241199-019-g10 HT0261 Homo sapiens cDNA
1371	13965	26491	1.86	3.4E-01	Y00554.1	NT	Azotobacter vinelandii nifA gene for NifA protein (positive regulatory element)
2445	15012	27584	2.6	3.4E-01	D90909.1	NT	Synechocystis sp. PCC6803 complete genome, 11/27, 1311235-1430418
3032	15648	28126	0.73	3.4E-01	AL163210.2	NT	Homo sapiens chromosome 21 segment HS21C010
3032	15648	28127	0.73	3.4E-01	AL163210.2	NT	Homo sapiens chromosome 21 segment HS21C010
3183	15796	28268	0.96	3.4E-01	D90909.1	NT	Synechocystis sp. PCC6803 complete genome, 11/27, 1311235-1430418
3197	15809	28282	6.78	3.4E-01	U83905.1	NT	Canis familiaris rod photoreceptor cGMP-gated channel alpha-subunit (CNGC1) mRNA, complete cds
3387	15996	28473	0.84	3.4E-01	AF034862.1	NT	Homo sapiens pulmonary surfactant protein D, promoter region and exon 1
3584	16188	28671	4.84	3.4E-01	AF106835.1	NT	Methylovorus sp. strain SS1 putative GrpE (grpE), DnaK (dnaK), and putative DnaJ (dnaJ) genes, complete cds
3858	16456		1.32	3.4E-01	BF449010.1	EST_HUMAN	7n94a01.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3572232 3' similar to TR:Q9UJ15 Q9UJ15 DJ18C9.1 ;
4108	16702		1.23	3.4E-01	AF184614.1	NT	Homo sapiens p47-phox (NCF1) gene, complete cds
4126	16718		1.56	3.4E-01	AA584196.1	EST_HUMAN	no11b10.s1 NCI_CGAP_Phe1 Homo sapiens cDNA clone IMAGE:1100347 3'
4594	17177	29624	0.7	3.4E-01	AF166341.1	NT	Homo sapiens integrin alpha 6 (ITGA6) gene, exons 12 through 23
4749	17330	29773	2	3.4E-01	BE069912.1	EST_HUMAN	MR4-BT0403-230200-202-c01 BT0403 Homo sapiens cDNA
4765	17346	29795	1.01	3.4E-01	BF314689.1	EST_HUMAN	601901632F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4130935 5'
5087	17660		4.2	3.4E-01	AI240973.1	EST_HUMAN	qj95c05.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867208 3' similar to contains Alu repetitive element;
5864	18486	31210	2.9	3.4E-01	AL161594.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 90
5979	18599		5.68	3.4E-01	AA085313.1	EST_HUMAN	zn12d11.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:547221 3'
6158	18771		2.44	3.4E-01	L02971.1	NT	Echovirus 22 1AB, 1C, 1D, 2A, 2B, 2C, 3A, 3B, 3C, 3D proteins RNA, complete mature peptides and cds

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6181	18791	31560	0.69	3.4E-01	BE748912.1	EST_HUMAN	601571811T1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838826 3'
6256	18865	31635	2.45	3.4E-01	AW204505.1	EST_HUMAN	UI-H-BI1-aei-e-12-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:2719582 3'
6382	18986	31766	1.62	3.4E-01	AL120544.1	EST_HUMAN	DKFZp761A249_r1 761 (synonym: hamy2) Homo sapiens cDNA clone DKFZp761A249 5'
6841	19431		1.32	3.4E-01	N95225.1	EST_HUMAN	zb53e12.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:307342 3'
7027	19561	32388	1.09	3.4E-01	AI468082.1	EST_HUMAN	tm63g05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2162840 3' similar to gb:S37431 LAMININ RECEPTOR (HUMAN);
7847	20389		0.51	3.4E-01	AE000493.1	NT	Escherichia coli K-12 MG1655 section 383 of 400 of the complete genome
8179	20720	33635	0.51	3.4E-01	Y14930.1	NT	Homo sapiens TCRAV28 gene, allele A4, partial
8424	20964		1.92	3.4E-01	AA337063.1	EST_HUMAN	EST41765 Endometrial tumor Homo sapiens cDNA 5' end
8498	21037	33958	0.88	3.4E-01	L04690.1	NT	Cricetulus griseus cholesterol 7-alpha-hydroxylase gene, complete cds
8786	21325	34249	1.63	3.4E-01	9633624	NT	Bovine enterovirus strain K2577, complete genome
9139	21674	34616	3.89	3.4E-01	P26013	SWISSPROT	INTEGRIN BETA-8 PRECURSOR
9139	21674	34617	3.89	3.4E-01	P26013	SWISSPROT	INTEGRIN BETA-8 PRECURSOR
9342	21856		0.59	3.4E-01	AB017510.1	NT	Ephydatia fluviatilis mRNA for PLC-gammaS, complete cds
9367	20306	33208	4.77	3.4E-01	U19492.1	NT	Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds
9367	20306	33209	4.77	3.4E-01	U19492.1	NT	Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds
9612	22112	35075	0.86	3.4E-01	U68763.1	NT	Glycine max putative transcription factor SCOF-1 (scof-1) mRNA, complete cds
9804	22302	35288	1.99	3.4E-01	AJ225084.1	NT	Homo sapiens FAA gene, exon 16, 17 and 18
10373	22867		0.54	3.4E-01	AE004096.1	NT	Vibrio cholerae chromosome I, section 4 of 251 of the complete chromosome
10895	23416		4.42	3.4E-01	AE000881.1	NT	Methanobacterium thermoautotrophicum from bases 1018444 to 1029212 (section 87 of 148) of the complete genome
10932	23450	36471	2.61	3.4E-01	P06925	SWISSPROT	PROBABLE E4 PROTEIN
10969	23484	36512	2.72	3.4E-01	AF045981.1	NT	Rutilus arcasii cytochrome b (cytb) gene, mitochondrial gene encoding mitochondrial protein, partial cds
11164	23671	36716	1.89	3.4E-01	M25856.1	NT	Human von Willebrand factor gene, exons 36 and 37
11164	23671	36717	1.89	3.4E-01	M25856.1	NT	Human von Willebrand factor gene, exons 36 and 37
11372	23824	36887	2.27	3.4E-01	AB035507.1	NT	Rattus norvegicus mRNA for a-glycerin/MUC18, complete cds
11401	23852	36917	4.36	3.4E-01	AL161515.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 27
11614	24056	37121	1.75	3.4E-01	BF061948.1	EST_HUMAN	7k69d12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3480646 3'
11655	24082		2.12	3.4E-01	U93604.1	NT	Citrus variegation virus putative replicase gene, partial cds
11771	24160		1.44	3.4E-01	Z21621.1	NT	S.cerevisiae RIB5 gene encoding Riboflavin synthase
11873	24836		1.8	3.4E-01	AF254351.1	NT	Schizosaccharomyces pombe Cwf8p (cwf8) gene, complete cds
11996	24303		14.59	3.4E-01	L26339.1	NT	Human autoantigen mRNA, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12023	24865		3.18	3.4E-01	BE218652.1	EST_HUMAN	hv42h08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3176127 3' similar to contains PTR5.t3 PTR5 repetitive element ;
12079	24974		2.44	3.4E-01	9838361	NT	Beta vulgaris mitochondrion, complete genome
12196	24424	30950	2.66	3.4E-01	AJ297131.1	NT	Mus musculus SIL, MAP_17, CYP_a, SCL & CYP_b genes
12427	25068		1.25	3.4E-01	AJ288948.1	NT	Clostridium cellulolyticum partial spoIVB gene and spo0A gene, strain ATCC 35319
12523	24839		2.55	3.4E-01	AF019413.1	NT	Homo sapiens HLA class III region containing tenascin X (tenascin-X) gene, partial cds; cytochrome P450 21-hydroxylase (CYP21B), complement component C4 (C4B) G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes,>
12651	24723		2.71	3.4E-01	11466174	NT	Naegleria gruberi mitochondrion, complete genome
16	12695	25151	13.68	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
110	12695	25151	3.75	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
473	13106	25599	1.08	3.3E-01	AL161545.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 45
661	13285	25788	1.87	3.3E-01	7662485	NT	Homo sapiens KIAA1100 protein (KIAA1100), mRNA
1242	13840	26358	2.96	3.3E-01	Q12446	SWISSPROT	PROLINE-RICH PROTEIN LAS17
1350	13945	26469	3.58	3.3E-01	BF568880.1	EST_HUMAN	602184016T1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300251 3'
1649	14241	26775	1.43	3.3E-01	6753685	NT	Mus musculus disintegrin 5 (Dtgn5), mRNA
1773	14363		1.44	3.3E-01	AA332734.1	EST_HUMAN	EST36722 Embryo, 8 week I Homo sapiens cDNA 5' end
2075	14655		1.22	3.3E-01	AF031148.1	NT	Methylococcus capsulatus strain Bath outer membrane protein MopB (mopB) gene, complete cds
2450	15017		5.41	3.3E-01	4507834	NT	Homo sapiens uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase) (UMPS) mRNA
2976	15592	28074	2.14	3.3E-01	AJ251805.1	NT	Bacteriophage phi-YeO3-12 complete genome
3049	15665		0.66	3.3E-01	O02743	SWISSPROT	INTERLEUKIN-12 ALPHA CHAIN PRECURSOR (IL-12A) (CYTOTOXIC LYMPHOCYTE MATURATION FACTOR 35 KD SUBUNIT) (CLMF P35)
3091	15706	28178	0.82	3.3E-01	AJ007932.2	NT	Streptomyces argillaceus mithramycin biosynthetic genes
3542	16147	28629	0.99	3.3E-01	AB012922.1	NT	Homo sapiens MTA1-L1 gene, complete cds
3882	16480	28942	2.14	3.3E-01	O84645	SWISSPROT	EXODEOXYRIBONUCLEASE V BETA CHAIN
3890	16489	28949	0.85	3.3E-01	P22602	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER COMPONENT PROTEINASE (HC-PRO); PROTEIN P3]
4037	16635	29104	1.49	3.3E-01	AL161498.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 10
4073	16669	29130	1.95	3.3E-01	AF200446.1	NT	Hypoxylon fragiforme chitin synthase gene, partial cds
4457	17043		1.44	3.3E-01	D31662.1	NT	Rattus norvegicus DNA for regucalcin, partial cds
4799	17377		1.57	3.3E-01	AI539114.1	EST_HUMAN	tp78b12.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2205407 3' similar to gb:X57522 ANTIGEN PEPTIDE TRANSPORTER 1 (HUMAN);
4964	17538	29980	1.44	3.3E-01	D64003.1	NT	Synechocystis sp. PCC6803 complete genome, 22/27, 2755703-2868766

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5527	18159	30573	2.75	3.3E-01	X89819.1	NT	R.norvegicus mRNA for 3'UTR of ubiquitin-like protein
5527	18159	30574	2.75	3.3E-01	X89819.1	NT	R.norvegicus mRNA for 3'UTR of ubiquitin-like protein
6101	18717	31468	1.72	3.3E-01	BE619650.1	EST_HUMAN	601472768T1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3875753 3'
6101	18717	31469	1.72	3.3E-01	BE619650.1	EST_HUMAN	601472768T1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3875753 3'
6188	18798	31567	12.71	3.3E-01	P05691	SWISSPROT	CIRCUMSPOROZOITE PROTEIN (CS)
6969	19546	32369	4.8	3.3E-01	AI628131.1	EST_HUMAN	ty84h01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285809 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;
6969	19546	32370	4.8	3.3E-01	AI628131.1	EST_HUMAN	ty84h01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285809 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;
7758	20266	33162	1.81	3.3E-01	N85146.1	EST_HUMAN	J2498F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone J2498 5' similar to TEGT
8497	21036	33957	17.55	3.3E-01	BF683954.1	EST_HUMAN	602140372F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301800 5'
8663	21202	34120	0.48	3.3E-01	BF210322.1	EST_HUMAN	601873281F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4097180 5'
8696	21235	34157	0.49	3.3E-01	AU126115.1	EST_HUMAN	AU126115 NT2RP1 Homo sapiens cDNA clone NT2RP1000130 5'
8696	21235	34158	0.49	3.3E-01	AU126115.1	EST_HUMAN	AU126115 NT2RP1 Homo sapiens cDNA clone NT2RP1000130 5'
9042	21579	34508	0.83	3.3E-01	Q62925	SWISSPROT	MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 1 (MAPK/ERK KINASE KINASE 1) (MEK KINASE 1) (MEKK 1)
9300	21900	34848	0.99	3.3E-01	BE828461.1	EST_HUMAN	CM3-ET0041-180500-187-d10 ET0041 Homo sapiens cDNA
9300	21900	34849	0.99	3.3E-01	BE828461.1	EST_HUMAN	CM3-ET0041-180500-187-d10 ET0041 Homo sapiens cDNA
9430	21939	34887	3.24	3.3E-01	N69866.1	EST_HUMAN	za67h01.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:297649 3'
9471	21870	34819	2.93	3.3E-01	BF376745.1	EST_HUMAN	RC4-TN0077-250800-011-g04 TN0077 Homo sapiens cDNA
9902	22399		2.16	3.3E-01	L41044.1	NT	Homo sapiens high-mobility group phosphoprotein (HMGI-C) gene, exons 1-3, complete cds
10600	23134	36147	2.93	3.3E-01	X63953.1	NT	D.mauritiana Adh gene
10600	23134	36148	2.93	3.3E-01	X63953.1	NT	D.mauritiana Adh gene
10905	23424		1.8	3.3E-01	BF526499.1	EST_HUMAN	602070802F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4213585 5'
11110	23620	36661	17.52	3.3E-01	BE219351.1	EST_HUMAN	hv51g02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3176978 3'
11223	23754	36812	4.97	3.3E-01	P47953	SWISSPROT	GALECTIN-3 (GALACTOSE-SPECIFIC LECTIN 3) (MAC-2 ANTIGEN) (IGE-BINDING PROTEIN) (35 KD LECTIN) (CARBOHYDRATE BINDING PROTEIN 35) (CBP 35) (LAMININ-BINDING PROTEIN) (LECTIN L-29) (CBP30)
11565	24012		4.68	3.3E-01	AA806621.1	EST_HUMAN	ob71g02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336850 3'
11584	12695	25151	2.48	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
11759	24153	36771	1.96	3.3E-01	6598319	NT	Homo sapiens aldehyde oxidase 1 (AOX1), mRNA
12510	24631		36.28	3.3E-01	AP000002.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position (2/7)
482	13115		1.79	3.2E-01	AF018261.1	NT	Rattus norvegicus EH domain binding protein Epsin mRNA, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
747	13367		0.76	3.2E-01	AL161561.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 61
1204	13804	26317	27.96	3.2E-01	AF047013.1	NT	Fusarium poae virus 1 RNA2 putative RNA dependent RNA polymerase gene, complete cds
1325	13919	26441	1.39	3.2E-01	Z50202.1	NT	P.vulgaris arc5-1 gene
1434	14027	26555	7.37	3.2E-01	Q48624	SWISSPROT	LACTOSE PERMEASE (LACTOSE-PROTON SYMPORT) (LACTOSE TRANSPORT PROTEIN)
1811	14401	26946	0.92	3.2E-01	Z36041.1	NT	S.cerevisiae chromosome II reading frame ORF YBR172c
1819	14409	26954	6.36	3.2E-01	AW957194.1	EST_HUMAN	EST369264 MAGE resequences, MAGD Homo sapiens cDNA
1819	14409	26955	6.36	3.2E-01	AW957194.1	EST_HUMAN	EST369264 MAGE resequences, MAGD Homo sapiens cDNA
1883	14469	27026	1.22	3.2E-01	AL111655.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
2205	14781	27354	2.89	3.2E-01	BF203817.1	EST_HUMAN	601868804F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4111512 5'
2578	15140		2.24	3.2E-01	7710079	NT	Mus musculus Pbx/knotted 1 homeobox (Pknox1), mRNA
2734	15289	27857	1.09	3.2E-01	AF060568.1	NT	Homo sapiens promyelocytic leukemia zinc finger protein (PLZF) gene, complete cds
3668	16269		0.78	3.2E-01	D10872.1	NT	Humam h NAT allele 3-2 gene for arylamine N-acetyltransferase
4023	16621		0.61	3.2E-01	AL161546.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 46
4483	17068	29518	1.64	3.2E-01	M18818.1	NT	Rabbit beta-like globin gene cluster encoding the epsilon, gamma, delta (pseudogene) and beta globin polypeptides, complete cds
4597	17181	29628	1.56	3.2E-01	Q10268	SWISSPROT	HYPOTHETICAL 81.7 KD PROTEIN C13G7.04C IN CHROMOSOME I PRECURSOR
4844	17422		6.52	3.2E-01	BF693617.1	EST_HUMAN	602081972F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4246505 5'
5009	17582	30025	0.63	3.2E-01	Q57081	SWISSPROT	CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 3 (CYTADHERENCE ACCESSORY PROTEIN 3) (ACCESSORY ADHESIN PROTEIN 3) (P69)
5174	17741	30170	0.58	3.2E-01	BE782748.1	EST_HUMAN	601465591F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3868799 5'
5392	17950	30383	0.93	3.2E-01	AY008847.1	NT	Homo sapiens interleukin 12 p40 subunit (IL12B) gene, IL12B-1 allele, complete cds
5476	18110	30519	2.5	3.2E-01	BE173964.1	EST_HUMAN	CM0-HT0569-060300-269-f10 HT0569 Homo sapiens cDNA
6112	18728	31481	1.18	3.2E-01	L27221.1	NT	Giardia intestinalis pyruvate:flavodoxin oxidoreductase and flanking genes
6831	19421		1.03	3.2E-01	AB002359.1	NT	Human mRNA for KIAA0361 gene, KIAA0361 protein
8113	20654	33563	1.33	3.2E-01	M60266.1	NT	Rat ISO-atrial natriuretic factor gene, complete cds
8210	20751	33665	0.51	3.2E-01	AJ231001.1	NT	Rattus norvegicus repeat; map NOS-D12Wox1
8308	20849	33772	11.34	3.2E-01	X02508.1	NT	H.sapiens gene fragment for acetylcholine receptor (AChR) alpha subunit exons 8, 9 and 3' flanking region
8311	20852	33777	16.78	3.2E-01	BF311635.1	EST_HUMAN	601897107F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126633 5'
8398	20938		1.43	3.2E-01	AL161574.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 70
8437	20977	33891	1.22	3.2E-01	BF246771.1	EST_HUMAN	601855580F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075627 5'
8437	20977	33892	1.22	3.2E-01	BF246771.1	EST_HUMAN	601855580F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075627 5'
8508	21047	33968	2.72	3.2E-01	AE002015.1	NT	Deinococcus radiodurans R1 section 152 of 229 of the complete chromosome 1
8605	21144	34058	0.69	3.2E-01	U51026.1	NT	Oryctolagus.cuniculus Ig H-chain pseudogene, V-region (VH6-a2) gene, partial cds

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8605	21144	34059	0.69	3.2E-01	U51026.1	NT	Oryctolagus cuniculus Ig H-chain pseudogene, V-region (VH6-a2) gene, partial cds
8997	21535	34465	0.58	3.2E-01	AL163204.2	NT	Homo sapiens chromosome 21 segment HS21C004
9007	21544		2.08	3.2E-01	M86511.1	NT	Human monocyte antigen CD14 (CD14) mRNA, complete cds
9077	21613	34547	0.48	3.2E-01	AF041829.1	NT	Homo sapiens 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PF2K) gene, exons 12 and 13
9077	21613	34548	0.48	3.2E-01	AF041829.1	NT	Homo sapiens 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PF2K) gene, exons 12 and 13
9905	22402	35376	3.22	3.2E-01	U44914.1	NT	Borrelia burgdorferi plasmid cp32-2, erpC and erpD genes, complete cds; and unknown genes
10100	22595	35588	0.51	3.2E-01	BE326230.1	EST_HUMAN	hv99f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3181569 3'
10213	22708		3.03	3.2E-01	AB011399.1	NT	Homo sapiens gene for AF-6, complete cds
10548	23084	36098	3.28	3.2E-01	T06813.1	EST_HUMAN	EST04702 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBDZ21
11796	24999		4.31	3.2E-01	L07288.1	NT	Drosophila melanogaster laminin A (Lam-A) mRNA, complete cds
12344	24525		4.65	3.2E-01	O83217	SWISSPROT	ELONGATION FACTOR TU (EF-TU)
12441	24819		3.37	3.2E-01	AF157625.1	NT	Bos taurus inositol 1,4,5-trisphosphate receptor type 1 mRNA, complete cds
12489	24618		1.94	3.2E-01	L39874.1	NT	Homo sapiens deoxycytidylate deaminase gene, complete cds
12550	25041	30504	1.98	3.2E-01	BE385776.1	EST_HUMAN	601275480F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3616746 5'
2695	15252	27823	2.39	3.1E-01	R18051.1	EST_HUMAN	ye90h06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:125051 5' similar to gb:M64241 QM PROTEIN (HUMAN);
2722	15403	27843	3.67	3.1E-01	7661971	NT	Homo sapiens KIAA0174 gene product (KIAA0174), mRNA
2722	15403	27844	3.67	3.1E-01	7661971	NT	Homo sapiens KIAA0174 gene product (KIAA0174), mRNA
2883	15501		1.35	3.1E-01	AW629036.1	EST_HUMAN	hi46h08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2975391 3'
3208	15820		3.53	3.1E-01	AB029069.1	NT	Mus musculus gene for Ser/Thr kinase KKIAMRE, exon 6
3978	16576	29046	0.91	3.1E-01	AJ251586.1	NT	Daucus carota mRNA for transcription factor E2F (E2F gene)
5055	17628	30072	0.79	3.1E-01	S68245.1	NT	carbonic anhydrase IV [rats, Sprague-Dawley, lung, mRNA, 1205 nt]
5097	17670	30109	0.82	3.1E-01	AE003984.1	NT	Xylella fastidiosa, section 130 of 229 of the complete genome
5206	17771	30194	0.98	3.1E-01	AL161503.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 15
5669	18296	30776	10.8	3.1E-01	AF176111.1	NT	Homo sapiens hepatocyte nuclear factor-3 alpha (HNF3A) gene, exon 1
5786	18411	31127	0.75	3.1E-01	Z74883.1	NT	S.cerevisiae chromosome XV reading frame ORF YOL141w
5796	18421		0.99	3.1E-01	Y13278.1	NT	Mus musculus mRNA for polycystin
5945	18565	31295	2.16	3.1E-01	AF184122.1	NT	Homo sapiens filamin 2 (FLN2) gene, exons 10 through 22
6592	19189	31992	2.63	3.1E-01	AW983549.1	EST_HUMAN	RC3-HN0001-310300-011-b04 HN0001 Homo sapiens cDNA
6654	19250	32052	0.87	3.1E-01	AI264458.1	EST_HUMAN	ql39d01.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1874689 3'
6784	19375	32191	0.81	3.1E-01	X71887.1	NT	H.sapiens gene for immunoglobulin kappa light chain variable region A8 and A9
6863	19597		0.9	3.1E-01	AW377354.1	EST_HUMAN	MR2-CT0222-281099-005-h05 CT0222 Homo sapiens cDNA

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Accession No.	Top Hit Database Source	Top Hit Descriptor
7048	24741	30458	2.4	3.1E-01	BE737392.1	EST_HUMAN	601306121F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3640420 5'
7871	20183	33071	1.18	3.1E-01	4885390	NT	Homo sapiens hyaluronan synthase 2 (HAS2), mRNA
8582	21121	34041	1.22	3.1E-01	R45318.1	EST_HUMAN	yg46f01.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:35639 3'
9816	22314	35295	0.45	3.1E-01	6679322	NT	Mus musculus phosphatidylinositol-4-phosphate 5-kinase, type 1 gamma (Pip5k1c), mRNA
9979	22474	35456	0.81	3.1E-01	BF696639.1	EST_HUMAN	602124743F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4281611 5'
9979	22474	35457	0.81	3.1E-01	BF696639.1	EST_HUMAN	602124743F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4281611 5'
10035	22530	35525	1.7	3.1E-01	AI244001.1	EST_HUMAN	qj81e11.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1863980 3' similar to gb:S55700 HYDROXYMETHYLGLUTARYL-COA LYASE PRECURSOR (HUMAN);
10204	22699		0.54	3.1E-01	T55325.1	EST_HUMAN	yb47h08.s1 Stratagene fetal spleen (#937205) Homo sapiens cDNA clone IMAGE:74367 3' similar to similar to gb:M91036_rna2 HEMOGLOBIN GAMMA-A AND GAMMA-G CHAINS (HUMAN)
10717	23245	36261	1.95	3.1E-01	BF216117.1	EST_HUMAN	601883592F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4095814 5'
11411	23862	36923	2.03	3.1E-01	7662291	NT	Homo sapiens KIAA0764 gene product (KIAA0764), mRNA
11921	24257		2.13	3.1E-01	AF294308.1	NT	Anolis opalinus isolate QS NADH dehydrogenase subunit 2 (ND2) gene, complete cds; mitochondrial gene for mitochondrial product
11960	24282		1.95	3.1E-01	AF304162.1	NT	Stizostedion vitreum 40S ribosomal protein S11 mRNA, partial cds
12108	24370		3.31	3.1E-01	AF195953.1	NT	Homo sapiens membrane-bound aminopeptidase P (XNPEP2) gene, complete cds
12496	24624		3.89	3.1E-01	AF196779.1	NT	Homo sapiens transcription factor IGHM enhancer 3, JM11 protein, JM4 protein, JM5 protein, T54 protein, JM10 protein, A4 differentiation-dependent protein, triple LIM domain protein 6, and synaptophysin genes, complete cds; and L-type calcium channel a>
12535	25035		1.62	3.1E-01	10946623	NT	Mus musculus peptidoglycan recognition protein-like (Pglyrpl-pending), mRNA
76	15382	25234	1.37	3.0E-01	6755083	NT	Mus musculus protein kinase C, epsilon (Pkce), mRNA
275	12932	25419	11.51	3.0E-01	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
1268	13863	26380	2.05	3.0E-01	AW300400.1	EST_HUMAN	xs63f08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2774343 3'
1555	14147	26680	6.64	3.0E-01	AJ006755.1	NT	Balaenoptera physalus gene encoding atrial natriuretic peptide
3248	15860		1.4	3.0E-01	AB030481.1	NT	Corynebacterium sp. ALY-1 alyPG gene for polyguluronate lyase, complete cds
3932	16530	28997	2.1	3.0E-01	AW817785.1	EST_HUMAN	PM1-ST0262-261199-001-g01 ST0262 Homo sapiens cDNA
4046	16643	29109	1.01	3.0E-01	AJ271736.1	NT	Homo sapiens Xq pseudoautosomal region; segment 2/2
4611	17194	29640	1.78	3.0E-01	AJ006755.1	NT	Balaenoptera physalus gene encoding atrial natriuretic peptide
5554	18186	30601	5.34	3.0E-01	BE741629.1	EST_HUMAN	601594960F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948734 5'
5627	18256	30726	0.88	3.0E-01	AF229247.1	NT	Cantagalo orthopoxvirus hemagglutinin gene, complete cds
5695	18321	30820	4.03	3.0E-01	BE693575.1	EST_HUMAN	RC3-BT0333-180700-111-a03 BT0333 Homo sapiens cDNA
5695	18321	30821	4.03	3.0E-01	BE693575.1	EST_HUMAN	RC3-BT0333-180700-111-a03 BT0333 Homo sapiens cDNA
5731	18357	31062	4.57	3.0E-01	U01247.1	NT	Mus musculus 129/sv Clara cell 10 kd protein (mCC10) gene, complete cds
6919	19578	32407	3.14	3.0E-01	D16313.1	NT	Mouse cytokeratin 15 gene, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6944	18052	30475	0.7	3.0E-01	U02369.1	NT	Strongylocentrotus purpuratus 34/67 kDa laminin-binding protein mRNA, partial cds
7005	19503	32322	0.96	3.0E-01	AF229247.1	NT	Cantagalo orthopoxvirus hemagglutinin gene, complete cds
7175	19707	32555	0.76	3.0E-01	AL163206.2	NT	Homo sapiens chromosome 21 segment HS21C006
7367	19893	32756	6	3.0E-01	10947007	NT	Mus musculus midnolin (Midn-pending), mRNA
7512	20033	32899	1.88	3.0E-01	AF071810.1	NT	Streptococcus pneumoniae strain DBL5 PspA (pspA) gene, partial cds
7867	20409	33316	1.07	3.0E-01	AE001755.1	NT	Thermotoga maritima section 67 of 136 of the complete genome
8314	20855		3.82	3.0E-01	9910161	NT	Mus musculus C-type (calcium dependent, carbohydrate recognition domain) lectin, superfamily member 9 (Clecsf9), mRNA
8411	20951	33870	1.27	3.0E-01	BE566083.1	EST_HUMAN	601339079F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3681594 5'
8763	21302	34223	0.82	3.0E-01	AF141676.1	NT	Streptomyces sulfonofaciens isopenicillin N synthase (pcbC) gene, partial cds
8805	21344		0.95	3.0E-01	7661685	NT	Homo sapiens DKFZP586M0122 protein (DKFZP586M0122), mRNA
9145	21680	34624	0.98	3.0E-01	AF220507.1	NT	Anabaena PCC7120 cytosine-specific DNA methyltransferase (dmnB) gene, complete cds; putative anthranilate phosphoribosyltransferase gene, partial cds; and unknown gene
9493	21993	34949	0.55	3.0E-01	P76389	SWISSPROT	HYPOTHETICAL 59.5 KD PROTEIN IN WZA-ASMA INTERGENIC REGION
9878	22375	35352	0.84	3.0E-01	BF574612.1	EST_HUMAN	602133271F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288336 5'
10294	22788	35778	0.56	3.0E-01	AW118111.1	EST_HUMAN	xs03d10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2606035 3'
10296	22790	35780	1.95	3.0E-01	AB030231.1	NT	Aspergillus oryzae bipA gene for ER chaperone BiP, complete cds
10315	22809	35801	0.73	3.0E-01	BF683841.1	EST_HUMAN	602140133F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301097 5'
10315	22809	35802	0.73	3.0E-01	BF683841.1	EST_HUMAN	602140133F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301097 5'
11604	24047	37112	2.87	3.0E-01	H51029.1	EST_HUMAN	yp84b10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:194107 5'
11604	24047	37113	2.87	3.0E-01	H51029.1	EST_HUMAN	yp84b10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:194107 5'
11975	24287		1.37	3.0E-01	P54660	SWISSPROT	PONTICULIN PRECURSOR
12227	24984		2.93	3.0E-01	AJ297631.1	NT	Rattus norvegicus mRNA for glyceraldehyde-3-phosphate dehydrogenase type 2 (gapdh-2 gene)
12529	25033		2.51	3.0E-01	6677766	NT	Mus musculus ribose 5-phosphate isomerase A (Rpia), mRNA
1924	14509	27064	2.27	2.9E-01	5174502	NT	Homo sapiens membrane component, chromosome 11, surface marker 1 (M11S1) mRNA
2070	14650	27221	1.38	2.9E-01	AE000736.1	NT	Aquifex aeolicus section 68 of 109 of the complete genome
2524	15088		1.22	2.9E-01	M32360.1	NT	Mouse apolipoprotein A-II (Alp-2) gene, complete cds
3289	15900	28379	2.07	2.9E-01	AW754239.1	EST_HUMAN	PM1-CT0326-171299-001-f12 CT0326 Homo sapiens cDNA
3289	15900	28380	2.07	2.9E-01	AW754239.1	EST_HUMAN	PM1-CT0326-171299-001-f12 CT0326 Homo sapiens cDNA
3965	16563	29032	0.71	2.9E-01	AI610836.1	EST_HUMAN	tp21a11.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188412 3' similar to gb:D15050 NIL-2-A ZINC FINGER PROTEIN (HUMAN);contains element L1 repetitive element ;
4159	16751		0.67	2.9E-01	AW002902.1	EST_HUMAN	wr02f10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2480395 3'
4583	17166	29609	1.21	2.9E-01	AA284468.1	EST_HUMAN	zs57d12.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:701591 5' similar to contains Alu repetitive element;

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4793	17372		0.63	2.9E-01	AL163207.2	NT	Homo sapiens chromosome 21 segment HS21C007
5148	17718		1.02	2.9E-01	U90756.1	NT	Lymantria dispar vitellogenin gene, complete cds
5154	17724	30155	1.43	2.9E-01	7662169	NT	Homo sapiens KIAA0537 gene product (KIAA0537), mRNA
5285	17847		1.7	2.9E-01	AI670899.1	EST_HUMAN	wa06f03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297309 3' similar to contains L1.t2 L1 repetitive element ;
5463	18098		1.25	2.9E-01	R37485.1	EST_HUMAN	yf77e12.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:28291 3'
5592	19522	32344	0.79	2.9E-01	AF321001.1	NT	Suaeda maritima subsp. salsa S-adenosylmethionine sythetase 2 mRNA, complete cds
5937	18558	31286	5.1	2.9E-01	X56098.1	NT	B.subtilis levanase operon levD, levE, levF, levG and sacC (partial) genes for fructose phosphotransferase system polypeptides P16,18,28,30 and levanase
5937	18558	31287	5.1	2.9E-01	X56098.1	NT	B.subtilis levanase operon levD, levE, levF, levG and sacC (partial) genes for fructose phosphotransferase system polypeptides P16,18,28,30 and levanase
5949	18570	31302	6.06	2.9E-01	6679682	NT	Mus musculus Eph receptor A8 (Epha8), mRNA
6206	18816	31587	1.26	2.9E-01	AA418145.1	EST_HUMAN	zv97b12.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:767711 5'
6427	19030	31813	0.93	2.9E-01	AI797128.1	EST_HUMAN	we27c05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2342312 3' similar to contains L1.t1 L1 repetitive element ;
6467	19068	31854	2.3	2.9E-01	U03420.1	NT	Bos taurus myosin I mRNA, complete cds
6585	19183	31983	0.72	2.9E-01	R69194.1	EST_HUMAN	yi39d08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:141615 5'
6585	19183	31984	0.72	2.9E-01	R69194.1	EST_HUMAN	yi39d08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:141615 5'
7062	18081	30437	1.35	2.9E-01	AF142329.1	NT	Mus musculus Fliih protein (Fliih) gene, complete cds; and Ligih protein (Ligih) gene, partial cds
7153	19685	32527	2.87	2.9E-01	Q04399	SWISSPROT	PUTATIVE MULTICOPPER OXIDASE YDR506C
7213	19744	32598	1.61	2.9E-01	AF100956.1	NT	Mus musculus major histocompatibility locus class II region; Fas-binding protein Daxx (DAXX) gene, partial cds; Bing1 (BING1), tapasin (tapasin), RalGDS-like factor (RLF), KE2 (KE2), BING4 (BING4), beta1, 3-galactosyl transferase (beta1,3-galactosyl tr>
7860	20402	33308	1.92	2.9E-01	BE540422.1	EST_HUMAN	601065830F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3452287 5'
7860	20402	33309	1.92	2.9E-01	BE540422.1	EST_HUMAN	601065830F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3452287 5'
8103	20644		0.89	2.9E-01	BF217743.1	EST_HUMAN	601882570F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4095113 5'
8531	21070	33990	0.66	2.9E-01	AU150910.1	EST_HUMAN	AU150910 NT2RP2 Homo sapiens cDNA clone NT2RP2003901 3'
8856	21395	34318	1.07	2.9E-01	AF225908.1	NT	Arabidopsis thaliana sulfonylurea receptor-like protein mRNA, complete cds
8963	21501	34423	0.71	2.9E-01	M22452.1	NT	Baboon lymphocyte homing/adhesion receptor mRNA, complete cds
9170	21747	34689	0.77	2.9E-01	AJ248287.1	NT	Pyrococcus abyssi complete genome; segment 5/6
9170	21747	34690	0.77	2.9E-01	AJ248287.1	NT	Pyrococcus abyssi complete genome; segment 5/6
10773	23297	36302	2.24	2.9E-01	AF128843.1	NT	Trypanosoma cruzi stage-specific surface glycoprotein gp82 (gp82) mRNA, partial cds
11041	23555	36589	2.88	2.9E-01	V01394.1	NT	Torpedo californica mRNA encoding acetylcholine receptor gamma subunit
11041	23555	36590	2.88	2.9E-01	V01394.1	NT	Torpedo californica mRNA encoding acetylcholine receptor gamma subunit

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11453	23903	36970	2.07	2.9E-01	AA935373.1	EST_HUMAN	ny35h02.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1273779 similar to contains LTR8.t2 LTR8 repetitive element ;
11456	23906	36973	5.52	2.9E-01	AL139078.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 5/6
12172	24411	30944	1.54	2.9E-01	AW005671.1	EST_HUMAN	wz88f05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2565921 3' similar to contains element MER29 repetitive element ;
12262	24472	30932	1.47	2.9E-01	AF092453.1	NT	Homo sapiens TNF-a-inducible RNA binding protein (TIRP) gene, complete cds
12313	24505		1.4	2.9E-01	BE788199.1	EST_HUMAN	601482059F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3884559 5'
12586	24679	30877	1.57	2.9E-01	Y08937.1	NT	Chlamydomonas reinhardtii mRNA for nitrite reductase structural locus
12586	24679	30878	1.57	2.9E-01	Y08937.1	NT	Chlamydomonas reinhardtii mRNA for nitrite reductase structural locus
594	13224		2.06	2.8E-01	U67136.1	NT	Rattus norvegicus A-kinase anchoring protein AKAP150 mRNA, complete cds
599	13228		0.75	2.8E-01	L28145.1	NT	Prune dwarf virus movement protein, complete cds; coat protein, complete cds
1122	13725	26238	3.14	2.8E-01	AF168050.1	NT	Guira guira oocyte maturation factor Mos (c-mos) gene, partial cds
1320	13914	26435	3.51	2.8E-01	BE313442.1	EST_HUMAN	601148733F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3163688 5'
1320	13914	26436	3.51	2.8E-01	BE313442.1	EST_HUMAN	601148733F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3163688 5'
1334	13928	26448	1.03	2.8E-01	D86550.1	NT	Human mRNA for serine/threonine protein kinase, complete cds
1765	14355	26902	2.01	2.8E-01	AW860020.1	EST_HUMAN	QV1-CT0364-120200-065-b05 CT0364 Homo sapiens cDNA
2057	14638	27210	2.12	2.8E-01	AL047620.1	EST_HUMAN	DKFZp586I2321_r1 586 (synonym: hute1) Homo sapiens cDNA clone DKFZp586I2321
2175	14752	27322	3.53	2.8E-01	AW511195.1	EST_HUMAN	hd44b03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2912333 3'
2511	15075	27648	2.41	2.8E-01	AE000494.1	NT	Escherichia coli K-12 MG1655 section 384 of 400 of the complete genome
2511	15075	27649	2.41	2.8E-01	AE000494.1	NT	Escherichia coli K-12 MG1655 section 384 of 400 of the complete genome
2584	15147		2.75	2.8E-01	AL161565.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 65
2688	15246	27813	1.21	2.8E-01	AB020975.1	NT	Arabidopsis thaliana mRNA for lipoyltransferase, complete cds
2988	15614		1.7	2.8E-01	AF179480.1	NT	Toxoplasma gondii 90kDa heat-shock protein (HSP90) mRNA, partial cds
2999	15615	28094	2.36	2.8E-01	Z14037.1	NT	B.taurus microsatellite (ETH121)
2999	15615	28095	2.36	2.8E-01	Z14037.1	NT	B.taurus microsatellite (ETH121)
3425	16033	28513	1.26	2.8E-01	AP000004.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 777001-994000 nt. position (4/7)
4068	16664	29125	2.06	2.8E-01	AE001180.1	NT	Borrelia burgdorferi (section 66 of 70) of the complete genome
4202	16791		0.62	2.8E-01	AE004450.1	NT	Pseudomonas aeruginosa PA01, section 11 of 529 of the complete genome
4276	16862		2.75	2.8E-01	AI090868.1	EST_HUMAN	ov44g10.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1640226 3' similar to contains Alu repetitive element;contains element MER22 repetitive element ;
4553	17136	29584	1.32	2.8E-01	AL021127.2	NT	Mus musculus chromosome X contigA; putative Magea9 gene, Caltractin, NAD(P) steroid dehydrogenase and Zinc finger protein 185
4559	17142	29590	2.47	2.8E-01	P13615	SWISSPROT	RNA POLYMERASE BETA SUBUNIT (LARGE STRUCTURAL PROTEIN) (L PROTEIN)
4897	17472	29928	1.07	2.8E-01	D15050.1	NT	Human mRNA for transcription factor AREB6, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4897	17472	29929	1.07	2.8E-01	D15050.1	NT	Human mRNA for transcription factor AREB6, complete cds
4937	17512	29958	1.02	2.8E-01	AW594539.1	EST_HUMAN	hg66d05.x1 NCI_CGAP_GC8 Homo sapiens cDNA clone IMAGE:2950569 3'
4949	17524	29965	1.17	2.8E-01	AF075238.1	NT	Hepatitis G virus isolate 60 (SZNAE12) polyprotein precursor, gene, partial cds
4955	17530	29972	3.5	2.8E-01	AF030154.1	NT	Bovine adenovirus 3 complete genome
4986	17560	30003	1.67	2.8E-01	BF528188.1	EST_HUMAN	602042601F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4180129 5'
5013	17587	30030	2.82	2.8E-01	AI272669.1	EST_HUMAN	qi59c11.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1876628 3' similar to contains Alu repetitive element;contains element LTR5 repetitive element ;
5514	24744	30558	21.59	2.8E-01	AA349997.1	EST_HUMAN	EST57072 Infant brain Homo sapiens cDNA 5' end
5791	18416	31132	2.76	2.8E-01	AB016625.1	NT	Homo sapiens OCTN2 gene, complete cds
5983	18603		0.97	2.8E-01	AW992583.1	EST_HUMAN	CM1-BN0024-150200-118-g12 BN0024 Homo sapiens cDNA
6078	18695	31442	0.69	2.8E-01	AA765296.1	EST_HUMAN	oa01d06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1303691 3' similar to gb:M34539 FK506-BINDING PROTEIN (HUMAN);
6323	25114		0.75	2.8E-01	M36688.1	NT	Bovine 680 bp repeated unit of 1.723 satellite DNA
6366	18970	31748	1.55	2.8E-01	AF003124.1	NT	Mesembryanthemum crystallinum fructose-biphosphate aldolase mRNA, complete cds
6366	18970	31749	1.55	2.8E-01	AF003124.1	NT	Mesembryanthemum crystallinum fructose-biphosphate aldolase mRNA, complete cds
6829	19419	32235	8.4	2.8E-01	BF511215.1	EST_HUMAN	UI-H-BI4-aoi-f-04-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone IMAGE:3085182 3'
7467	19989		1.19	2.8E-01	U05633.1	NT	Marsilea quadrifolia ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, chloroplast gene encoding chloroplast protein, partial cds
7768	20276	33174	0.69	2.8E-01	BE537151.1	EST_HUMAN	601063105F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3449569 5'
8036	20578	33482	1.12	2.8E-01	AI346126.1	EST_HUMAN	qp48h01.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1926289 3' similar to gb:X06323_cds1 MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (HUMAN);
8036	20578	33483	1.12	2.8E-01	AI346126.1	EST_HUMAN	qp48h01.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1926289 3' similar to gb:X06323_cds1 MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L3 (HUMAN);
8150	20691	33605	2.16	2.8E-01	U51688.1	NT	Homo sapiens lanosterol 14-alpha demethylase cytochrome P450 (CYP51) gene, exon 5
8451	20991	33909	0.47	2.8E-01	AA911629.1	EST_HUMAN	of02h05.s1 NCI_CGAP_Co12 Homo sapiens cDNA clone IMAGE:1419993 3' similar to gb:M87789 IG GAMMA-1 CHAIN C REGION (HUMAN);
8525	21064		6.69	2.8E-01	BF347847.1	EST_HUMAN	602022987F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4156525 5'
9387	21810	34761	1.22	2.8E-01	U17251.1	NT	Neurospora crassa negative regulator sulfur controller-2 (scon-2) gene, complete cds
9627	22127		1.03	2.8E-01	L13654.1	NT	Lycopersicon esculentum peroxidase (TPX1) mRNA, complete cds
9803	22301	35286	1.04	2.8E-01	AF132728.1	NT	Escherichia coli translocated intimin receptor Tir (tir) gene, complete cds
9803	22301	35287	1.04	2.8E-01	AF132728.1	NT	Escherichia coli translocated intimin receptor Tir (tir) gene, complete cds
9881	22358	35338	0.84	2.8E-01	AF294393.1	NT	Rattus norvegicus glycerol-3-phosphate dehydrogenase gene, promoters A and B and exons 1a and 1b; nuclear gene for mitochondrial product
9972	22467	35451	1.91	2.8E-01	7706163	NT	Homo sapiens hypothetical protein (LOC51319), mRNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10215	22710		0.81	2.8E-01	9626154	NT	Fujinami sarcoma virus, complete genome
10255	22750	35738	0.47	2.8E-01	BE959727.2	EST_HUMAN	601654822R1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:3839765 3'
10622	23154	36166	2.26	2.8E-01	BF241062.1	EST_HUMAN	601880794F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4109350 5'
10622	23154	36167	2.26	2.8E-01	BF241062.1	EST_HUMAN	601880794F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4109350 5'
10651	23183	36197	2.83	2.8E-01	BF695970.1	EST_HUMAN	601852148F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4076026 5'
10760	23284	36297	3.31	2.8E-01	AF051662.1	NT	Drosophila heteroneura fruitless (fru) gene, alternative splice products, 5' flanking region, exons 1 through 7 and complete cds
11158	23665		4.56	2.8E-01	BF674023.1	EST_HUMAN	602137418F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4273853 5'
12213	24436		15.74	2.8E-01	D83329.1	NT	Mus musculus DNA for prostaglandin D2 synthase, complete cds
12328	24514	30920	8.89	2.8E-01	BE178699.1	EST_HUMAN	PM4-HT0606-030400-001-e07 HT0606 Homo sapiens cDNA
12356	24533	30927	1.25	2.8E-01	BE900116.1	EST_HUMAN	601673020F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3955996 5'
12519	24988		2.21	2.8E-01	11433629	NT	Homo sapiens CDC42-binding protein kinase beta (DMPK-like) (CDC42BPB), mRNA
502	13134	25622	3.21	2.7E-01	Y17324.1	NT	Rattus norvegicus CDK104 mRNA
641	13264	25740	2.53	2.7E-01	AA450061.1	EST_HUMAN	zx39b10.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:788827 3' similar to contains Alu repetitive element;
1304	13898	26418	1.69	2.7E-01	AB004906.1	NT	Ipomoea purpurea transposable element Tip100 gene for transposase, complete cds
1662	14255		2.17	2.7E-01	X79815.1	NT	G.lamblia SR2 gene
1767	14357	26903	3.34	2.7E-01	W58067.1	EST_HUMAN	zd22h10.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:341443 5'
1813	14403	26948	4.14	2.7E-01	P03341	SWISSPROT	GAG POLYPROTEIN [CONTAINS: INNER COAT PROTEIN P12; CORE PROTEIN P15; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]
2181	15459		2.77	2.7E-01	AF047575.1	NT	Rattus norvegicus vesicular monoamine transporter type 2, promoter region and exon 1
2405	14973	27545	7.35	2.7E-01	Y13868.1	NT	Feline immunodeficiency virus env gene, isolate ITTO088PIU (M88), partial
2496	15060	27634	3.82	2.7E-01	AI310858.1	EST_HUMAN	ta43c11.x2 NCI_CGAP_Lu25 Homo sapiens cDNA clone IMAGE:2046836 3' similar to contains element L1 repetitive element ;
3013	15629		0.73	2.7E-01	BF088284.1	EST_HUMAN	CM1-HT0875-080900-385-e05 HT0875 Homo sapiens cDNA
4082	16678	29138	1.98	2.7E-01	AI928015.1	EST_HUMAN	wo92e11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2462828 3'
4096	16691	29147	0.79	2.7E-01	AF216214.1	NT	Drosophila buzzatii alpha-esterase 6 (aE6) gene, partial cds
4096	16691	29148	0.79	2.7E-01	AF216214.1	NT	Drosophila buzzatii alpha-esterase 6 (aE6) gene, partial cds
4101	16695	29151	2.31	2.7E-01	L77569.1	NT	Homo sapiens DiGeorge syndrome critical region, telomeric end
5020	17594	30037	0.98	2.7E-01	L27516.1	NT	Triticum aestivum (Wcs66) gene, complete cds
5193	17758		3.82	2.7E-01	AW856131.1	EST_HUMAN	RC1-CT0286-230200-016-e03 CT0286 Homo sapiens cDNA
5471	18105	30424	2.07	2.7E-01	P17277	SWISSPROT	HOMEOBOX PROTEIN HOX-A4 (CHOX-1.4)
5681	18308		0.96	2.7E-01	AB033171.1	NT	Astreopora myriophthalma mitochondrial cytb gene for cytochrome b, partial cds
6724	19318	32122	1.07	2.7E-01	AE001094.1	NT	Archaeoglobus fulgidus section 13 of 172 of the complete genome

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6724	19318	32123	1.07	2.7E-01	AE001094.1	NT	Archaeoglobus fulgidus section 13 of 172 of the complete genome
6875	19609	32443	2.03	2.7E-01	Q61554	SWISSPROT	FIBRILLIN 1 PRECURSOR
7085	19656	32495	0.76	2.7E-01	U15967.1	NT	Drosophila melanogaster rfc40 protein, Rop protein (Rop), and small GTP binding protein (DRes2) genes, complete cds
7393	19918	32782	0.87	2.7E-01	Q11079	SWISSPROT	HYPOTHETICAL 20.9 KD PROTEIN B0563.3 IN CHROMOSOME X
7562	20079	32954	0.95	2.7E-01	Q01168	SWISSPROT	NITROGEN REGULATORY PROTEIN NUT1
7562	20079	32955	0.95	2.7E-01	Q01168	SWISSPROT	NITROGEN REGULATORY PROTEIN NUT1
7677	20188	33076	2.21	2.7E-01	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
7677	20188	33077	2.21	2.7E-01	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
7720	20228	33116	0.92	2.7E-01	AA351121.1	EST_HUMAN	EST58740 Infant brain Homo sapiens cDNA 5' end similar to similar to myosin-binding protein H
7720	20228	33117	0.92	2.7E-01	AA351121.1	EST_HUMAN	EST58740 Infant brain Homo sapiens cDNA 5' end similar to similar to myosin-binding protein H
7805	20348	33256	0.95	2.7E-01	AA013147.1	EST_HUMAN	ze35b11.s1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:360957 3' similar to contains Alu repetitive element;
7969	20511		0.51	2.7E-01	AF048820.1	NT	Carassius auratus pituitary adenylate cyclase activating polypeptide type 1 receptor precursor mRNA, complete cds
8079	20621	33534	0.59	2.7E-01	AW868503.1	EST_HUMAN	MR1-SN0062-100500-002-d09 SN0062 Homo sapiens cDNA
8127	20668	33577	0.48	2.7E-01	R39257.1	EST_HUMAN	yc91h06.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:23511 3'
8232	20773	33694	0.94	2.7E-01	AL161552.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 52
8694	21233	34154	0.83	2.7E-01	Q14764	SWISSPROT	MAJOR VAULT PROTEIN (MVP) (LUNG RESISTANCE-RELATED PROTEIN)
8957	21495	34418	0.53	2.7E-01	X03216.1	NT	Staphylococcus aureus transposon Tn554
9256	21782	34734	9.93	2.7E-01	O83809	SWISSPROT	THREONYL-TRNA SYNTHETASE (THREONINE--TRNA LIGASE) (THRRS)
9256	21782	34735	9.93	2.7E-01	O83809	SWISSPROT	THREONYL-TRNA SYNTHETASE (THREONINE--TRNA LIGASE) (THRRS)
9259	21785		2.02	2.7E-01	P37928	SWISSPROT	FIMBRIAE W PROTEIN
9716	22214	35188	0.67	2.7E-01	D89660.1	NT	Rattus norvegicus DNA for peroxisome assembly factor-2, exon 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16, 17 and complete cds
9993	22488	35476	0.91	2.7E-01	AF091848.1	NT	Oryctolagus cuniculus calgranulin C mRNA, partial cds
10026	22521	35517	2.5	2.7E-01	AF087434.1	NT	Mus musculus transcription factor NF-ATc isoform a (NF-ATca) mRNA, complete cds
10148	22643	35634	0.69	2.7E-01	AF156539.1	NT	Homo sapiens xeroderma pigmentosum complementation group C (XPC) gene, intron 9
10148	22643	35635	0.69	2.7E-01	AF156539.1	NT	Homo sapiens xeroderma pigmentosum complementation group C (XPC) gene, intron 9
10691	23221	36233	2.31	2.7E-01	AV705043.1	EST_HUMAN	AV705043 ADB Homo sapiens cDNA clone ADBCOD05 5'
10691	23221	36234	2.31	2.7E-01	AV705043.1	EST_HUMAN	AV705043 ADB Homo sapiens cDNA clone ADBCOD05 5'

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10702	23231	36244	3.65	2.7E-01	AJ133269.1	NT	Homo sapiens caveolin-1/-2 locus, Contig1, D7S522, genes CAV2 (exons 1, 2a, and 2b), CAV1 (exons 1 and 2)
12296	24863		1.72	2.7E-01	AB008782.1	NT	Arabidopsis thaliana mRNA for sulfate transporter, complete cds
12501	24627		3.96	2.7E-01	AF217491.1	NT	Homo sapiens fragile 16D oxido reductase (FOR) gene, exon 6
495	15416	25615	2.06	2.6E-01	P78411	SWISSPROT	IROQUOIS-CLASS HOMEODOMAIN PROTEIN IRX-2
506	13139		1.24	2.6E-01	D16459.1	NT	Bos taurus mRNA for mb-1, complete cds
1437	14030	26558	2.19	2.6E-01	BE885087.1	EST_HUMAN	601510838F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912345 5'
1485	14078	26616	1.36	2.6E-01	AB013290.1	NT	Glycine max pseudogene for Bd 30K
1941	14525	27080	6.59	2.6E-01	AL161472.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2
1941	14525	27081	6.59	2.6E-01	AL161472.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2
2140	14718		13.12	2.6E-01	AW733152.1	EST_HUMAN	bb04d10.x1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:2958451 3' similar to gb:M36072 60S RIBOSOMAL PROTEIN L7A (HUMAN); gb:M14689_cds1 Mouse surfeit locus surfeit 3 protein gene (MOUSE);
2197	14773	27347	1.41	2.6E-01	M11844.1	NT	Human prealbumin gene, complete cds
2512	15076		2.09	2.6E-01	Y12996.1	NT	B.maritimus rbcL gene
2583	15146		10.77	2.6E-01	BE272440.1	EST_HUMAN	601126016F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:2990043 5'
3640	16243	28719	0.86	2.6E-01	M22342.1	NT	Bacteriphage T2 DNA-(adenine-N6)methyltransferase (dam) gene, complete cds
3705	16306	28775	2.13	2.6E-01	AF229118.1	NT	Homo sapiens acetylcholinesterase collagen-like tail subunit (COLQ) gene, exons 1A, 2, 3, 4, and 5
4175	16766	29214	0.7	2.6E-01	AW959510.1	EST_HUMAN	EST371580 MAGE resequences, MAGF Homo sapiens cDNA
4234	16822	29273	19.98	2.6E-01	BE080598.1	EST_HUMAN	QV1-BT0830-040400-132-e03 BT0830 Homo sapiens cDNA
4448	17034	29477	1.57	2.6E-01	AF175293.1	NT	Enterococcus faecium strain N97-330 vanD glycopeptide resistance gene cluster, complete cds; and unknown gene
4593	17176	29622	0.76	2.6E-01	AB021180.1	NT	Gallus gallus mRNA for skeletal myosin heavy chain, complete cds
4593	17176	29623	0.76	2.6E-01	AB021180.1	NT	Gallus gallus mRNA for skeletal myosin heavy chain, complete cds
4646	17228	29684	1.35	2.6E-01	AA457617.1	EST_HUMAN	aa89d07.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:838477 5'
4752	17333	29776	1.63	2.6E-01	U01103.1	NT	Arabidopsis thaliana PSI type III chlorophyll a/b-binding protein (Lhca3*1) mRNA, complete cds
4825	17403	29856	1.47	2.6E-01	AF142703.1	NT	Ophrestia radicosa maturase-like protein (matK) gene, complete cds; chloroplast gene for chloroplast product
5107	17679	30118	3.56	2.6E-01	H04858.1	EST_HUMAN	yj51e05.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:152288 5'
5195	17760		0.58	2.6E-01	AA884625.1	EST_HUMAN	am33b11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1468605 3'
5544	18176		1.29	2.6E-01	AB035972.1	NT	Paramecium caudatum gene for PAP, complete cds
5640	18269	30742	0.68	2.6E-01	M96060.1	NT	Acetobacter xylinum cellulose synthase (bcsA) gene, partial cds, CMCax and CcpAx genes, complete cds

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5760	18388		0.81	2.6E-01	AI862398.1	EST_HUMAN	td16a03.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2075788 3' similar to contains element MER35 repetitive element ;
5947	18568	31299	0.69	2.6E-01	AF207550.1	NT	Homo sapiens protein translocase, JM26 protein, UDP-galactose translocator, pim-2 protooncogene homolog pim-2h, and shal-type potassium channel genes, complete cds; JM12 protein and transcription factor IGHM enhancer 3 genes, partial cds; and unknown g>
6221	25113		2.36	2.6E-01	AE001811.1	NT	Thermotoga maritima section 123 of 136 of the complete genome
6348	18953	31732	1.89	2.6E-01	AI582557.1	EST_HUMAN	ts02e12.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2227438 3' similar to SW:NDF1_RAT Q64289 NEUROGENIC DIFFERENTIATION FACTOR 1 ;contains element LTR1 repetitive element ;
6348	18953	31733	1.89	2.6E-01	AI582557.1	EST_HUMAN	ts02e12.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2227438 3' similar to SW:NDF1_RAT Q64289 NEUROGENIC DIFFERENTIATION FACTOR 1 ;contains element LTR1 repetitive element ;
6554	19152	31948	1.05	2.6E-01	AL162757.2	NT	Neisseria meningitidis serogroup A strain Z2491 complete genome; segment 6/7
7103	19673	32512	0.97	2.6E-01	AI914380.1	EST_HUMAN	wd48c04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2331366 3' similar to gb:M37721 PEPTIDYL-GLYCINE ALPHA-AMIDATING MONOOXYGENASE PRECURSOR (HUMAN);
7457	24783		0.96	2.6E-01	AL139077.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 4/6
7721	20229	33118	1.6	2.6E-01	R10365.1	EST_HUMAN	yf37a03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:129004 3' similar to gb:X12517 U1 SMALL NUCLEAR RIBONUCLEOPROTEIN C (HUMAN);
7791	20334	33240	1.14	2.6E-01	R02411.1	EST_HUMAN	ye82a07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:124212 5'
7845	20387	33290	1.18	2.6E-01	BE144331.1	EST_HUMAN	MR0-HT0166-181199-003-d12 HT0166 Homo sapiens cDNA
8083	20625	33538	0.67	2.6E-01	X82641.1	NT	D.melanogaster mRNA for alpha 1,2 mannosidase (Berlin)
8083	20625	33539	0.67	2.6E-01	X82641.1	NT	D.melanogaster mRNA for alpha 1,2 mannosidase (Berlin)
8276	20817	33738	2.99	2.6E-01	BF343588.1	EST_HUMAN	602014422F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4150396 5'
8349	20890	33810	1.89	2.6E-01	Q10199	SWISSPROT	HYPOTHETICAL 75.2 KD PROTEIN C11C11.02 IN CHROMOSOME II
8627	21166	34080	4.49	2.6E-01	BE830339.1	EST_HUMAN	RC5-ET0082-310500-021-F10 ET0082 Homo sapiens cDNA
8627	21166	34081	4.49	2.6E-01	BE830339.1	EST_HUMAN	RC5-ET0082-310500-021-F10 ET0082 Homo sapiens cDNA
9388	21811	34762	0.96	2.6E-01	X17604.1	NT	S. occidentalis INV gene for invertase (EC 3.2.1.26)
9654	22153		0.5	2.6E-01	AF057121.1	NT	Lontra canadensis cytochrome b (cytb) gene, mitochondrial gene encoding mitochondrial protein, complete cds
9782	22280	35265	0.93	2.6E-01	P87366	SWISSPROT	GREEN-SENSITIVE OPSIN (GREEN CONE PHOTORECEPTOR PIGMENT) (KFH-G)
9782	22280	35266	0.93	2.6E-01	P87366	SWISSPROT	GREEN-SENSITIVE OPSIN (GREEN CONE PHOTORECEPTOR PIGMENT) (KFH-G)
10093	22588		0.5	2.6E-01	Q28295	SWISSPROT	VON WILLEBRAND FACTOR PRECURSOR (VWF)
10403	22897		0.91	2.6E-01	Y10196.1	NT	Homo sapiens PHEX gene
10500	22994		0.51	2.6E-01	AI978681.1	EST_HUMAN	wr58b09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2491865 3'

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11300	23752	36809	2.18	2.6E-01	P48280	SWISSPROT	CELL DIVISION PROTEIN FTSW HOMOLOG
11400	23851		30.68	2.6E-01	X51755.1	NT	Human lambda-immunoglobulin constant region complex (germline)
11777	24165		1.72	2.6E-01	10190655	NT	Mus musculus jerky (Jrk), mRNA
11973	24991		4.06	2.6E-01	BE883491.1	EST_HUMAN	601511052F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912612 5'
12042	24329	30996	4.8	2.6E-01	AF316896.1	NT	Homo sapiens Na/K-ATPase gamma subunit (FXYD2) gene, complete cds, alternatively spliced
12396	24556		1.34	2.6E-01	D88425.1	NT	Cavia cobaya mRNA for serine/threoine kinase, complete cds
12478	24612		1.96	2.6E-01	AE001713.1	NT	Thermotoga maritima section 25 of 136 of the complete genome
12526	24641		1.37	2.6E-01	AF141325.2	NT	Homo sapiens inositol polyphosphate 1-phosphatase (INPP1) gene, complete cds
12567	24666		3.74	2.6E-01	P47285	SWISSPROT	HYPOTHETICAL PROTEIN MG039
262	12921	25407	1.48	2.5E-01	4502296	NT	Homo sapiens ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit (ATP5D), nuclear gene encoding mitochondrial protein, mRNA
263	12921	25407	1.77	2.5E-01	4502296	NT	Homo sapiens ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit (ATP5D), nuclear gene encoding mitochondrial protein, mRNA
276	12933		4.29	2.5E-01	M26501.1	NT	Starfish (P.ochraceus) cytoplasmic actin gene, complete cds
865	13480	25994	1.02	2.5E-01	U09964.1	NT	Mus musculus ICR/Swiss glyceraldehyde 3-phosphate dehydrogenase (Gapd-S) gene, complete cds
1098	13703		1.03	2.5E-01	AE002156.1	NT	Ureaplasma urealyticum section 57 of 59 of the complete genome
1160	13763	26274	11.59	2.5E-01	T89837.1	EST_HUMAN	ye11g07.r1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:117468 5'
1566	14158	26689	0.87	2.5E-01	AL115624.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
1766	14356		6.09	2.5E-01	4885406	NT	Homo sapiens hyperpolarization activated cyclic nucleotide-gated potassium channel 4 (HCN4) mRNA
1927	15454	27067	1.29	2.5E-01	BE696604.1	EST_HUMAN	PM4-CT0400-310700-005-d08 CT0400 Homo sapiens cDNA
1927	15454	27068	1.29	2.5E-01	BE696604.1	EST_HUMAN	PM4-CT0400-310700-005-d08 CT0400 Homo sapiens cDNA
2452	15019		12.93	2.5E-01	AE000675.1	NT	Aquifex aeolicus section 7 of 109 of the complete genome
2536	15100	27673	0.93	2.5E-01	6679216	NT	Mus musculus protein-L-isoaspartate (D-aspartate) O-methyltransferase 1 (Pcmt1), mRNA
2540	15104		1.49	2.5E-01	AA251987.1	EST_HUMAN	zs11a12.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:684862 5'
3459	16066		3.41	2.5E-01	AW973471.1	EST_HUMAN	EST385464 MAGE resequences, MAGM Homo sapiens cDNA
3587	16191	28675	0.84	2.5E-01	AF233875.1	NT	Danio rerio peptide YY precursor gene, complete cds
3603	16207	28685	7.97	2.5E-01	AL161517.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 29
4143	16735		1.36	2.5E-01	P32323	SWISSPROT	A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR
4409	16994		0.9	2.5E-01	Q03314	SWISSPROT	RHIB PROTEIN
4722	17303	29747	0.59	2.5E-01	AF242431.1	NT	Mus musculus neuronal apoptosis inhibitory protein 6 (Naip6) gene, complete cds; and Naip3 gene, exons 2-9 and 11-16
4860	17438		1.47	2.5E-01	Q27225	SWISSPROT	MOLT-INHIBITING HORMONE PRECURSOR (MIH)

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4869	17445	29896	4.69	2.5E-01	AF007768.1	NT	Choristoneura fumiferana diapause associated protein 2 (DAP2) mRNA, complete cds
4896	17471	29927	2.82	2.5E-01	AE004416.1	NT	Vibrio cholerae chromosome II, section 73 of 93 of the complete chromosome
4924	17499		3.21	2.5E-01	AJ230113.1	NT	Mus musculus annexin V gene, intron 4 segment containing 5' LTR and gag portion of MuERV-L (murine endogenous retrovirus) element
4959	17534	29976	0.61	2.5E-01	BE896785.1	EST_HUMAN	601437468F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922600 5'
5262	12933		0.65	2.5E-01	M26501.1	NT	Starfish (P.ochraceus) cytoplasmic actin gene, complete cds
5529	18161	30576	12.86	2.5E-01	S83390.1	NT	T3 receptor-associating cofactor-1 [human, fetal liver, mRNA, 2930 nt]
6114	18730		0.84	2.5E-01	AL163207.2	NT	Homo sapiens chromosome 21 segment HS21C007
6738	19332	32138	0.83	2.5E-01	AJ251973.1	NT	Homo sapiens partial steerin-1 gene
7389	19914	32778	0.82	2.5E-01	U13992.1	NT	Feline calicivirus CFI/68 RNA helicase/cysteine protease/RNA-dependent RNA polymerase polyprotein precursor and capsid protein precursor, genes, complete cds; and unknown gene
7413	19938		1.35	2.5E-01	AF134119.1	NT	Mus musculus SKD1 (Skd1) gene, complete cds
7632	20144	33025	4.48	2.5E-01	AL163282.2	NT	Homo sapiens chromosome 21 segment HS21C082
7786	20329	33236	2.31	2.5E-01	BF109040.1	EST_HUMAN	7157e03.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3525389 3'
7797	20340	33248	0.7	2.5E-01	BE960712.1	EST_HUMAN	601653391R2 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3826198 3'
8168	20709	33625	2.2	2.5E-01	BF038595.1	EST_HUMAN	601459238F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3862809 5'
8336	20877	33798	0.72	2.5E-01	P04492	SWISSPROT	E1B PROTEIN, SMALL T-ANTIGEN (E1B 19K)
8571	21110	34029	3.03	2.5E-01	H53236.1	EST_HUMAN	yq84f07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:202501 5'
8808	21347	34271	0.88	2.5E-01	M88626.1	NT	Mouse testis-specific protein (TPX-1) gene, exon 10
9435	21961	34909	15.98	2.5E-01	U89651.2	NT	Homo sapiens matrix metalloproteinase MMP Rasi-1 gene, promoter region
9435	21961	34910	15.98	2.5E-01	U89651.2	NT	Homo sapiens matrix metalloproteinase MMP Rasi-1 gene, promoter region
9492	21948	34897	2.09	2.5E-01	AF085164.1	NT	Hordeum vulgare receptor-like kinase LRK10 gene, partial cds
9492	21948	34898	2.09	2.5E-01	AF085164.1	NT	Hordeum vulgare receptor-like kinase LRK10 gene, partial cds
10010	22505	35496	1.66	2.5E-01	AW581997.1	EST_HUMAN	RC3-ST0186-130100-015-a07 ST0186 Homo sapiens cDNA
10436	22930	35937	1.53	2.5E-01	AW152246.1	EST_HUMAN	xq40c10.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2630034 3' similar to contains Alu repetitive element;contains element MSR1 repetitive element ;
10439	22933	35941	1.31	2.5E-01	X58491.1	NT	Mouse L1Md LINE DNA
10459	22953	35962	2.03	2.5E-01	AL161505.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 17
10459	22953	35963	2.03	2.5E-01	AL161505.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 17
10955	23470	36495	4.3	2.5E-01	D50914.1	NT	Human mRNA for KIAA0124 gene, partial cds
11712	24122	37153	5.29	2.5E-01	AF200528.1	NT	Zea mays cellulose synthase-4 (CesA-4) mRNA, complete cds
11740	25075		10.13	2.5E-01	AL161541.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 41
579	13209	25687	1.67	2.4E-01	AA936316.1	EST_HUMAN	on70d04.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1562023 3'
881	13495	26014	2.4	2.4E-01	BF576124.1	EST_HUMAN	602132442F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271578 5'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1347	13942	26464	21.36	2.4E-01	AJ289880.1	NT	Homo sapiens KIAA0851 gene (partial), XT3 gene and LZTFL1 gene
1347	13942	26465	21.36	2.4E-01	AJ289880.1	NT	Homo sapiens KIAA0851 gene (partial), XT3 gene and LZTFL1 gene
1427	14020	26548	0.93	2.4E-01	Y17293.1	NT	Homo sapiens FLI-1 gene, partial
1891	14476		27.27	2.4E-01	AF267753.1	NT	Mesembryanthemum crystallinum putative potassium channel protein Mkt1p mRNA, complete cds
1944	14528	27084	1.17	2.4E-01	AF251708.1	NT	Zaocys dhumnades fructose-1,6-bisphosphatase mRNA, complete cds
2079	14659	27230	1.49	2.4E-01	AI742958.1	EST_HUMAN	wg76d05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2371017 3' similar to TR:O60267 O60267 KIAA0512 PROTEIN. ;
2183	14759	27329	1.04	2.4E-01	AF111168.2	NT	Homo sapiens serine palmitoyl transferase, subunit II gene, complete cds; and unknown genes
2213	14788		1.04	2.4E-01	P45384	SWISSPROT	IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (IGA1 PROTEASE)
2302	14875	27451	1.78	2.4E-01	AE000680.1	NT	Aquifex aeolicus section 12 of 109 of the complete genome
2425	14993	27566	1.26	2.4E-01	BF002171.1	EST_HUMAN	7h23d04.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:3316807 3' similar to SW:PRSB_XENLA O42586 26S PROTEASE REGULATORY SUBUNIT 6A ;
2575	15138	27708	3.05	2.4E-01	Z36534.1	NT	D.discoideum (Ax3-K) ponA gene
2790	15343	27913	1.79	2.4E-01	X71783.1	NT	S.pombe swi6 gene
2812	15364	27933	3.88	2.4E-01	AF030154.1	NT	Bovine adenovirus 3 complete genome
3166	15780		3.27	2.4E-01	U72726.1	NT	Oryza longistaminata receptor kinase-like protein, family member D, and retrofit (gag/pol) genes, complete cds
3182	15795	28267	1.38	2.4E-01	X74209.1	NT	H.sapiens AGT gene, PstI fragment of intron 4
3724	16325	28792	1.26	2.4E-01	AF169793.1	NT	Podospora anserina HET-C protein (Het-c) gene, complete cds
3824	16424	28886	0.83	2.4E-01	AE000312.1	NT	Escherichia coli K-12 MG1655 section 202 of 400 of the complete genome
4103	16697		0.6	2.4E-01	D29960.1	NT	Rattus norvegicus mRNA for alphaB crystallin-related protein, complete cds
5008	17581	30024	1.08	2.4E-01	AE000305.1	NT	Escherichia coli K-12 MG1655 section 195 of 400 of the complete genome
5220	17785	30203	0.93	2.4E-01	BE737592.1	EST_HUMAN	601572862F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:3839775 5'
5302	17864		1.55	2.4E-01	K02402.1	NT	Human coagulation factor IX gene, complete cds
5653	18280	30758	0.83	2.4E-01	AI925707.1	EST_HUMAN	wo33d05.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2457129 3'
5653	18280	30759	0.83	2.4E-01	AI925707.1	EST_HUMAN	wo33d05.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2457129 3'
5676	18303	30785	0.85	2.4E-01	D50871.1	NT	Glycine max mRNA for mitotic cyclin b1-type, complete cds
5836	18460	31182	7.92	2.4E-01	AF091216.1	NT	Mus musculus Wrn protein (Wrn) gene, complete cds
5836	18460	31183	7.92	2.4E-01	AF091216.1	NT	Mus musculus Wrn protein (Wrn) gene, complete cds
6050	24754		1.02	2.4E-01	AJ133836.2	NT	Branchiostoma floridae mRNA for calmodulin 2 (caM2 gene)
6054	18672	31411	2.36	2.4E-01	BF592336.1	EST_HUMAN	7i54d04.x1 NCI_CGAP_Br16 Homo sapiens cDNA clone IMAGE:3338503 3' similar to SW:SFR4_HUMAN Q08170 SPLICING FACTOR, ARGININE/SERINE-RICH 4 ;contains element TAR1 TAR1 repetitive element ;
6138	18752	31510	2.5	2.4E-01	AF035546.1	NT	Drosophila melanogaster p38a MAP kinase gene, complete cds

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6240	18849	31619	2.26	2.4E-01	7661801	NT	Homo sapiens HSPC142 protein (HSPC142), mRNA
6290	18898	31669	0.8	2.4E-01	AV733787.1	EST_HUMAN	AV733787 cdA Homo sapiens cDNA clone cdAADE11 5'
6656	19252	32055	2.43	2.4E-01	AI698989.1	EST_HUMAN	wc62c11.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2323220 3' similar to gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN PRECURSOR (HUMAN);
7381	19907	32772	8.84	2.4E-01	L43001.1	NT	Bos taurus guanylyl cyclase-activating protein 2 (guca2) mRNA, complete cds
7709	20218	33106	1.06	2.4E-01	AF229644.1	NT	Mus musculus DXImx48e protein (DXImx48e) mRNA, complete cds
8139	20680	33591	0.71	2.4E-01	AJ006397.1	NT	Streptococcus pneumoniae rr08 and hk08 genes; two component system 08
8139	20680	33592	0.71	2.4E-01	AJ006397.1	NT	Streptococcus pneumoniae rr08 and hk08 genes; two component system 08
8290	20831	33752	1.66	2.4E-01	AJ012585.1	NT	Tetrahymena thermophila macronuclear gene encoding ribosomal protein L3, exons 1-2
8535	21074	33994	0.97	2.4E-01	BF242794.1	EST_HUMAN	601877679F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4106298 5'
8588	21127		0.58	2.4E-01	BF678275.1	EST_HUMAN	602086188F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4250372 5'
9059	21596	34526	0.58	2.4E-01	AL139077.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 4/6
9059	21596	34527	0.58	2.4E-01	AL139077.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 4/6
9482	21881	34826	6.84	2.4E-01	AI693515.1	EST_HUMAN	wd43e02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2330906 3' similar to contains MER22.b1 TAR1 repetitive element ;
9620	22120	35083	0.6	2.4E-01	AF220067.1	NT	Drosophila melanogaster SKPB gene, complete cds
9620	22120	35084	0.6	2.4E-01	AF220067.1	NT	Drosophila melanogaster SKPB gene, complete cds
10335	22829	35823	1.95	2.4E-01	Q03692	SWISSPROT	COLLAGEN ALPHA 1(X) CHAIN PRECURSOR
10647	23179	36192	3.25	2.4E-01	AL161494.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 6
10715	23243	36260	2.9	2.4E-01	AF030199.1	NT	Mus musculus type 1 sigma receptor gene, complete cds
11081	23593		2.28	2.4E-01	Z21647.1	NT	P.asiatica mosaic virus genomic RNA
11665	24089	37145	1.91	2.4E-01	AF217491.1	NT	Homo sapiens fragile 16D oxido reductase (FOR) gene, exon 6
11807	24853		2.65	2.4E-01	AF004213.1	NT	Arabidopsis thaliana ethylene-insensitive3-like1 (EIL1) mRNA, complete cds
11866	24220		2.02	2.4E-01	AJ278191.1	NT	Mus musculus mRNA for putative mc7 protein (mc7 gene)
12086	24838		2.18	2.4E-01	V01507.1	NT	Gallus gallus gene coding for a-actin
12320	25061		1.5	2.4E-01	BF229975.1	EST_HUMAN	RC3-CT0413-100800-023-b06 CT0413 Homo sapiens cDNA
12562	24682		2.31	2.4E-01	AL163281.2	NT	Homo sapiens chromosome 21 segment HS21C081
412	13047	25538	0.91	2.3E-01	S75898.1	NT	aromatase [Poephila guttata=zebra finches, ovary, mRNA, 3188 nt]
665	13289		4.4	2.3E-01	U39713.1	NT	Mycoplasma genitalium section 35 of 51 of the complete genome
695	13318	25803	17.02	2.3E-01	U67596.1	NT	Methanococcus jannaschii section 138 of 150 of the complete genome
969	13580	26092	3.44	2.3E-01	BE311893.1	EST_HUMAN	601142073F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3505818 5'
1647	14239	26774	1.19	2.3E-01	AJ245480.1	NT	Brassica napus slg gene for S-locus glycoprotein, cultivar T2
1674	14266	26800	2.75	2.3E-01	Y10887.2	NT	Mus musculus cdh5 gene, exon 1, partial
2089	14669		1.29	2.3E-01	AJ235353.1	NT	Homo sapiens partial intron 3 of the wild type AF-4/FEL gene

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2489	15054	27626	2.03	2.3E-01	BE297718.1	EST_HUMAN	601175562F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3531015 5'
2678	15236	27803	1.16	2.3E-01	M11319.1	NT	Human erythropoietin gene, complete cds
2851	14024	26552	1.42	2.3E-01	AB015033.1	NT	Marinilabilia agarovorans gyrB gene for DNA gyrase subunit B, partial cds, strain:IFO 14957
2990	15606	28086	0.93	2.3E-01	AA601379.1	EST_HUMAN	no16d06.s1 NCI_CGAP_Phe1 Homo sapiens cDNA clone IMAGE:1100843 3' similar to contains Alu repetitive element;contains element THR repetitive element ;
3120	15734		6.96	2.3E-01	R21732.1	EST_HUMAN	yh21b07.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:130357 3'
3417	16025	28507	0.78	2.3E-01	H69836.1	EST_HUMAN	yr97h10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:213283 5'
3908	16507	28969	1.02	2.3E-01	S82821.1	NT	GSTA5=glutathione S-transferase Yc2 subunit {5' region, intron 1} [rats, Morris hepatoma cell line, Genomic, 2212 nt, segment 1 of 3]
4009	16607		5.14	2.3E-01	7662133	NT	Homo sapiens KIAA0450 gene product (KIAA0450), mRNA
4442	17028	29468	0.83	2.3E-01	R82252.1	EST_HUMAN	yj17f01.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:149017 5'
4489	17074		2.4	2.3E-01	L78789.1	NT	Mus musculus renin (Ren-1c) gene, promoter region
4548	17131	29578	0.87	2.3E-01	D90899.1	NT	Synechocystis sp. PCC6803 complete genome, 1/27, 1-133859
4586	17169	29613	2.16	2.3E-01	AF092535.1	NT	Homo sapiens mitogen-activated protein kinase p38delta (PRKM13) mRNA, complete cds
4652	17234	29690	6.13	2.3E-01	5031984	NT	Homo sapiens nuclear transport factor 2 (placental protein 15) (PP15) mRNA
5180	17585	30028	0.62	2.3E-01	J03280.1	NT	Human phenylethanolamine N-methyltransferase gene, complete cds
5202	17767	30191	0.62	2.3E-01	AB032400.1	NT	Mus musculus tulip 1 mRNA, complete cds
5403	17961	30372	0.9	2.3E-01	AE000240.1	NT	Escherichia coli K-12 MG1655 section 130 of 400 of the complete genome
5507	18140	30552	2.39	2.3E-01	AB040945.1	NT	Homo sapiens mRNA for KIAA1512 protein, partial cds
5621	18250	30718	2.05	2.3E-01	BF058381.1	EST_HUMAN	7k30b06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3476699 3' similar to SW:GAG_SMSAV P03330 GAG POLYPROTEIN [CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]. ;
5721	18347	31050	4.56	2.3E-01	X96587.1	NT	C.familiaris rom1 gene
5831	18455		1.19	2.3E-01	L39112.1	NT	Vittaforma corneum small subunit ribosomal RNA gene
5926	18548	31274	0.78	2.3E-01	S60371.1	NT	23S rRNA [Leuconostoc carnosum, Genomic, 2866 nt]
6096	18712	31461	2.34	2.3E-01	AI708840.1	EST_HUMAN	as27e12.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone IMAGE:2318446 3' similar to gb:X13238 CYTOCHROME C OXIDASE POLYPEPTIDE VIC PRECURSOR (HUMAN);
6096	18712	31462	2.34	2.3E-01	AI708840.1	EST_HUMAN	as27e12.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone IMAGE:2318446 3' similar to gb:X13238 CYTOCHROME C OXIDASE POLYPEPTIDE VIC PRECURSOR (HUMAN);
6762	19355	32164	0.76	2.3E-01	AF198089.1	NT	Oryctolagus cuniculus cytochrome oxidase subunit VIa (coxVIa2) mRNA, complete cds; nuclear gene for mitochondrial product
6959	19536	32360	4.1	2.3E-01	AI718148.1	EST_HUMAN	as42f12.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone IMAGE:2319887 3' similar to contains Alu repetitive element;
7165	19697	32544	0.7	2.3E-01	8923323	NT	Homo sapiens hypothetical protein FLJ20345 (FLJ20345), mRNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7331	19858	32721	0.69	2.3E-01	AF000227.1	NT	Secale cereale omega secalin gene, complete cds
7445	19969	32837	2.42	2.3E-01	AF175389.1	NT	Glycine max resistance protein LM17 precursor RNA, partial cds
7603	20116		3.63	2.3E-01	6754779	NT	Mus musculus myosin XV (Myo15), mRNA
7608	20121	32998	1.63	2.3E-01	BE888071.1	EST_HUMAN	601511573F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912859 5'
7732	20240		2.88	2.3E-01	N80983.1	EST_HUMAN	za12e08.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:292358 5'
7793	20336	33243	0.58	2.3E-01	AL161558.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 58
7935	20477	33387	1.93	2.3E-01	M68931.1	NT	Oxytricha nova macronuclear telomere-binding protein alpha subunit (tel-alpha alanine version) gene, complete cds
8430	20970	33882	0.6	2.3E-01	U57999.1	NT	Mus musculus prosaposin (psap\SGP-1) gene, complete cds
9087	21604	34534	0.87	2.3E-01	AA372164.1	EST_HUMAN	EST84061 Rhabdomyosarcoma Homo sapiens cDNA 5' end similar to DnaJ homolog (GB:X63368)
9087	21604	34535	0.87	2.3E-01	AA372164.1	EST_HUMAN	EST84061 Rhabdomyosarcoma Homo sapiens cDNA 5' end similar to DnaJ homolog (GB:X63368)
9501	22001	34958	0.65	2.3E-01	6679318	NT	Mus musculus phosphatidylinositol 3-kinase catalytic subunit delta (Pik3cd), mRNA
9644	22144	35112	0.51	2.3E-01	BE277860.1	EST_HUMAN	601120110F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2966739 5'
9697	22196	35169	0.76	2.3E-01	AW964460.1	EST_HUMAN	EST376533 MAGE resequences, MAGH Homo sapiens cDNA
9748	22244	35225	1.22	2.3E-01	X52124.1	NT	Haemophilus influenzae genes for HincII restriction-modification system (HincII methyltransferase (EC 2.1.1.72) and HincII endonuclease (EC 3.1.21.4))
9781	22279	35264	0.55	2.3E-01	AW364633.1	EST_HUMAN	PM2-DT0036-281299-001-f04 DT0036 Homo sapiens cDNA
9847	22345	35326	2.6	2.3E-01	BE173060.1	EST_HUMAN	MR0-HT0559-240400-014-g11 HT0559 Homo sapiens cDNA
9903	22400	35373	1.93	2.3E-01	AJ293261.1	NT	Rhizobium leguminosarum partial genomic DNA for exopolysaccharide biosynthesis genes
10339	22833	35828	0.94	2.3E-01	AF201929.1	NT	Murine hepatitis virus strain 2, complete genome
10351	22845		5.86	2.3E-01	BF133577.1	EST_HUMAN	601646155R2 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4102092 3'
10893	23414	36432	1.85	2.3E-01	AF004833.1	NT	Mus musculus tissue factor pathway inhibitor (TFPI) mRNA, complete cds
10893	23414	36433	1.85	2.3E-01	AF004833.1	NT	Mus musculus tissue factor pathway inhibitor (TFPI) mRNA, complete cds
11068	23580	36619	1.85	2.3E-01	AJ250189.1	NT	Mus musculus partial mRNA for muscle protein 534 (mg534 gene)
11068	23580	36620	1.85	2.3E-01	AJ250189.1	NT	Mus musculus partial mRNA for muscle protein 534 (mg534 gene)
11230	23761	36817	2.49	2.3E-01	AE002187.2	NT	Chlamydophila pneumoniae AR39, section 4 of 94 of the complete genome
11624	24066		1.6	2.3E-01	AV709736.1	EST_HUMAN	AV709736 ADC Homo sapiens cDNA clone ADCAGH01 5'
11788	24172		2.82	2.3E-01	U45426.1	NT	Borrelia burgdorferi 2.9-6 locus, ORF-A-D genes, complete cds and REP+ gene, partial cds
11876	24226		57.94	2.3E-01	T27231.1	EST_HUMAN	HCOEST44 HT29M6 Homo sapiens cDNA clone HCoE44 5'
11899	24804		1.31	2.3E-01	AA089819.1	EST_HUMAN	chn1424.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5'
11908	24246		1.61	2.3E-01	AW863940.1	EST_HUMAN	PM4-SN0012-030400-001-b06 SN0012 Homo sapiens cDNA
11969	25002	30610	3.1	2.3E-01	AW303623.1	EST_HUMAN	xv21d07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2813773 3' similar to TR:Q9Z175 Q9Z175 LYSYL OXIDASE-RELATED PROTEIN 2. ;contains PTR5.b2 TAR1 repetitive element ;

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed In Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12007	25053	30511	10.96	2.3E-01	BE882464.1	EST_HUMAN	601507202F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3908689 5'
12057	24340		1.94	2.3E-01	BF663319.1	EST_HUMAN	602144459F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4297719 5'
12107	24369		3.11	2.3E-01	AJ006519.1	NT	Rattus norvegicus mRNA for acid gated ion channel
12205	24429		1.36	2.3E-01	U49845.1	NT	Pleurodeles waltl distal-less like protein PwDlx-3 (PwDlx-3) mRNA, complete cds
12211	24369		1.67	2.3E-01	AJ006519.1	NT	Rattus norvegicus mRNA for acid gated ion channel
12480	24614		2.57	2.3E-01	BF475611.1	EST_HUMAN	nac39h12.x1 Lupski_sciatic_nerve Homo sapiens cDNA clone IMAGE:3395950 3' similar to contains element MER38 repetitive element ;
12668	24888	30710	1.26	2.3E-01	AA094108.1	EST_HUMAN	cl1864.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5'
93	12769	25252	0.91	2.2E-01	AI052190.1	EST_HUMAN	oz14a10.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1675290 3' similar to TR:Q13040 Q13040 ATP-BINDING CASSETTE PROTEIN ;
1611	14204	26738	2.85	2.2E-01	AF187850.1	NT	Homo sapiens PPAR delta gene, promoter region
2063	14643		3.89	2.2E-01	AF171901.1	NT	Trimeresurus malabaricus cytb gene, partial cds; mitochondrial gene for mitochondrial product
2136	14714	27287	3.16	2.2E-01	M34640.1	NT	Fresh-water sponge Emf1 alpha collagen (COLF1) gene
2447	15014	27586	5.61	2.2E-01	BF677538.1	EST_HUMAN	602085608F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249969 5'
2623	15185	27751	1.27	2.2E-01	BE618258.1	EST_HUMAN	601462629F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866190 5'
2623	15185	27752	1.27	2.2E-01	BE618258.1	EST_HUMAN	601462629F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866190 5'
2703	15260		1.17	2.2E-01	AL163218.2	NT	Homo sapiens chromosome 21 segment HS21C018
2906	15523	27993	4.28	2.2E-01	BE155625.1	EST_HUMAN	PM2-HT0353-281299-003-a12 HT0353 Homo sapiens cDNA
2906	15523	27994	4.28	2.2E-01	BE155625.1	EST_HUMAN	PM2-HT0353-281299-003-a12 HT0353 Homo sapiens cDNA
2947	15563		1.64	2.2E-01	AF020503.1	NT	Homo sapiens FRA3B common fragile region, diadenosine triphosphate hydrolase (FHIT) gene, exon 5
3439	16047		2.67	2.2E-01	AL161562.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 62
3886	16484		1.18	2.2E-01	AF155728.1	NT	Xiphophorus maculatus truncated Rex1 retrotransposon reverse transcriptase (RT) pseudogene
4291	16877		1.26	2.2E-01	AF119102.1	NT	Drosophila melanogaster UNC-119 (unc-119) gene, complete cds
4300	16886	29330	7.03	2.2E-01	AF155142.1	NT	Mus musculus mixed lineage kinase 3 (Mlk3) and two pore domain K+ channel subunit (Kcnk6) genes, complete cds
4350	16937	29378	2.59	2.2E-01	AF117340.1	NT	Mus musculus MAP kinase kinase kinase 1 (Mekk1) mRNA, complete cds
4350	16937	29379	2.59	2.2E-01	AF117340.1	NT	Mus musculus MAP kinase kinase kinase 1 (Mekk1) mRNA, complete cds
4447	17033	29475	1.36	2.2E-01	U01307.1	NT	Human scRNA (BC200 beta) pseudogene
4447	17033	29476	1.36	2.2E-01	U01307.1	NT	Human scRNA (BC200 beta) pseudogene
4952	17527		1.35	2.2E-01	D50604.1	NT	Human beta-cytoplasmic actin (ACTBP9) pseudogene
4957	17532	29974	2.86	2.2E-01	AA211216.1	EST_HUMAN	zq87c05.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:648968 5'
5196	17761		1.33	2.2E-01	L13299.1	NT	Mus musculus vinculin gene, exon 3
5203	17768		1.79	2.2E-01	AE001137.1	NT	Borrelia burgdorferi (section 23 of 70) of the complete genome

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5291	17853	30277	1.2	2.2E-01	BE141035.1	EST_HUMAN	MR0-HT0067-201099-002-c10 HT0067 Homo sapiens cDNA
5316	17878		0.9	2.2E-01	S57565.1	NT	histamine H2-receptor [rats, Genomic, 1928 nt]
5919	18541	31267	2.46	2.2E-01	5803002	NT	Homo sapiens diaphanous (Drosophila, homolog) 2 (DIAPH2), transcript variant 156, mRNA
5930	18552		3.53	2.2E-01	D64000.1	NT	Synechocystis sp. PCC6803 complete genome, 19/27, 2392729-2538999
6150	18763	31525	0.73	2.2E-01	U67087.1	NT	Gallus gallus T-box containing protein (Ch-TbxT) mRNA, complete cds
6150	18763	31526	0.73	2.2E-01	U67087.1	NT	Gallus gallus T-box containing protein (Ch-TbxT) mRNA, complete cds
6807	19398	32212	0.85	2.2E-01	AB038490.1	NT	Homo sapiens gene for fukutin, complete cds
7093	19684	32503	9.14	2.2E-01	AV756238.1	EST_HUMAN	AV756238 BM Homo sapiens cDNA clone BMFAHC06 5'
7183	19715	32562	1.46	2.2E-01	AF082738.1	NT	Streptococcus pyogenes phosphotidylglycerophosphate synthase (pgsA) and ABC transporter ATP-binding protein (stpA) genes, complete cds; and unknown genes
7183	19715	32563	1.46	2.2E-01	AF082738.1	NT	Streptococcus pyogenes phosphotidylglycerophosphate synthase (pgsA) and ABC transporter ATP-binding protein (stpA) genes, complete cds; and unknown genes
7333	19860	32723	2.01	2.2E-01	M24136.1	NT	Human glycophorin B gene, exon 4
7333	19860	32724	2.01	2.2E-01	M24136.1	NT	Human glycophorin B gene, exon 4
7688	20197	33085	0.68	2.2E-01	AF287967.1	NT	Homo sapiens homeobox B7 (HOXB7) gene, partial cds; and homeobox B6 (HOXB6), homeobox B5 (HOXB5), homeobox B4 (HOXB4), and homeobox B3 (HOXB3) genes, complete cds
7963	20505		3.06	2.2E-01	AF155143.1	NT	Mus musculus nm23-M1 gene, promoter region
8032	20574	33479	0.84	2.2E-01	Z49933.1	NT	E.coli sepA and sepB genes
8815	21354	34277	0.57	2.2E-01	L23312.1	NT	Mouse HD protein mRNA, complete cds
8815	21354	34278	0.57	2.2E-01	L23312.1	NT	Mouse HD protein mRNA, complete cds
8827	21366	34290	3.48	2.2E-01	AE001713.1	NT	Thermotoga maritima section 25 of 136 of the complete genome
8847	21386	34310	1.02	2.2E-01	U09964.1	NT	Mus mucculus ICR/Swiss glyceraldehyde 3-phosphate dehydrogenase (Gapd-S) gene, complete cds
8952	21490		3.12	2.2E-01	AW855039.1	EST_HUMAN	PM3-CT0263-241299-009-b07 CT0263 Homo sapiens cDNA
9043	21580	34509	1.82	2.2E-01	8393247	NT	Mus mucculus deformed epidermal autoregulatory factor 1 (Drosophila) (Deaf1), mRNA
9126	21661	34604	1.95	2.2E-01	BF376354.1	EST_HUMAN	MR1-TN0045-110900-006-c02 TN0045 Homo sapiens cDNA
9213	21730	34673	1.24	2.2E-01	W02988.1	EST_HUMAN	za04f08.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:291591 5'
9231	21953	34903	14.03	2.2E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
9274	21800	34750	0.74	2.2E-01	AJ009839.1	NT	Xenopus laevis mRNA for kinesin-like protein 3 (xklp3)
9285	21885	34830	0.71	2.2E-01	7657428	NT	Mus musculus osteoblast specific factor 2 (OSF-2), mRNA
9298	21898	34845	3.69	2.2E-01	M89643.1	NT	Brachydanio rerio ependymin beta and gamma chains (Epd) gene, complete cds
9539	22039	35000	0.59	2.2E-01	Q90980	SWISSPROT	CYCLIC NUCLEOTIDE GATED CHANNEL, ROD PHOTORECEPTOR, ALPHA SUBUNIT (CNG CHANNEL 3) (CNG-3) (CNG3)

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9729	22227	35204	3.1	2.2E-01	AF197941.1	NT	Funaria hygrometrica chloroplast-localized small heat shock protein (CPsHSP21) mRNA, complete cds; nuclear gene for chloroplast product
9864	22361	35341	2.23	2.2E-01	BF206507.1	EST_HUMAN	601869724F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4100189 5'
10079	22574	35569	0.87	2.2E-01	9625671	NT	Human herpesvirus 5, complete genome
10340	22834		0.61	2.2E-01	AF071001.1	NT	Mus musculus PHR1 (Phr1) gene, partial cds
10384	22878	35870	0.72	2.2E-01	AE001562.1	NT	Helicobacter pylori, strain J99 section 123 of 132 of the complete genome
10384	22878	35871	0.72	2.2E-01	AE001562.1	NT	Helicobacter pylori, strain J99 section 123 of 132 of the complete genome
11005	23519	36554	1.6	2.2E-01	AF257772.1	NT	Homo sapiens RNA binding protein MCG10 gene, complete cds, alternatively spliced
11299	23751	36808	5.56	2.2E-01	X01918.1	NT	Drosophila 68C glue gene cluster
11335	23033	36042	3.18	2.2E-01	7706215	NT	Homo sapiens H-2K binding factor-2 (LOC51580), mRNA
11715	24125		1.8	2.2E-01	BE870959.1	EST_HUMAN	601446957F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850670 5'
11827	25065		6.34	2.2E-01	U82671.2	NT	Homo sapiens chromosome Xq28 melanoma antigen family A2a (MAGEA2A), melanoma antigen family A12 (MAGEA12), melanoma antigen family A2b (MAGEA2B), melanoma antigen family A3 (MAGEA3), caltractin (CALT), NAD(P)H dehydrogenase-like protein (NSDHL), and LI>
11910	24248		5.37	2.2E-01	AF188843.1	NT	Vitis vinifera cultivar Pinot Noir plasma membrane aquaporin (PIP1a) mRNA, complete cds
12024	18029	30491	1.7	2.2E-01	AW361098.1	EST_HUMAN	RC1-CT0249-141199-021-g04 CT0249 Homo sapiens cDNA
12025	24317		1.85	2.2E-01	AW661922.1	EST_HUMAN	hi17b02.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2972523 3'
12575	25058		4.05	2.2E-01	AV694801.1	EST_HUMAN	AV694801 GKC Homo sapiens cDNA clone GKCAHB02 5'
12659	24730	30855	2.44	2.2E-01	BF243095.1	EST_HUMAN	601876452F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4104996 5'
1006	13617	26132	1.36	2.1E-01	AA569289.1	EST_HUMAN	nm31e11.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:1061804
1009	13619	26134	1.27	2.1E-01	AL161504.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 16
1163	13765		2.41	2.1E-01	AE002314.2	NT	Chlamydia muridarum, section 45 of 85 of the complete genome
1240	13838	26354	0.85	2.1E-01	6754299	NT	Mus musculus interferon (alpha and beta) receptor 2 (Ifnar2), mRNA
1240	13838	26355	0.85	2.1E-01	6754299	NT	Mus musculus interferon (alpha and beta) receptor 2 (Ifnar2), mRNA
1557	14149	26681	3.45	2.1E-01	AJ249895.1	NT	Mus musculus mas proto-oncogene and Igf2r gene for insulin-like growth factor type 2 and L41ps and Au76 pseudogenes
1957	14541	27097	1.84	2.1E-01	AA906824.1	EST_HUMAN	ok73e02.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1519610 3' similar to gb:K02765 COMPLEMENT C3 PRECURSOR (HUMAN);
2201	14777	27350	3.39	2.1E-01	BF695073.1	EST_HUMAN	602083129F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4247503 5'
2385	14954	27526	2.01	2.1E-01	6753235	NT	Mus musculus calcium channel, voltage dependent, alpha2/delta subunit 3 (Cacna2d3), mRNA
2951	15567	28041	2.53	2.1E-01	6912445	NT	Homo sapiens potassium voltage-gated channel, subfamily H (eag-related), member 4 (KCNH4), mRNA
3879	16477		6.58	2.1E-01	9838361	NT	Beta vulgaris mitochondrion, complete genome
4129	16721	29176	1.22	2.1E-01	P11675	SWISSPROT	IMMEDIATE-EARLY PROTEIN IE180

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4129	16721	29177	1.22	2.1E-01	P11675	SWISSPROT	IMMEDIATE-EARLY PROTEIN IE180
4336	16923		1.38	2.1E-01	AF124526.1	NT	Orchestia cavimana calcium-binding protein BP23 precursor (BP23) gene, complete cds
4465	17051		1.51	2.1E-01	AB033041.1	NT	Homo sapiens mRNA for KIAA1215 protein, partial cds
4876	17258	29709	1.83	2.1E-01	AB010273.1	NT	Homo sapiens pshsp47 gene, complete cds
5083	17656	30097	1.63	2.1E-01	U76409.1	NT	Lycopersicon esculentum homeobox 1 protein (THox1) mRNA, partial cds
5434	17990	30396	0.98	2.1E-01	J05082.1	NT	Vampire bat (D.rotundus) plasminogen activator mRNA, complete cds
5504	18138	30548	6.55	2.1E-01	BF672695.1	EST_HUMAN	602152001F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293001 5'
6987	19544	32368	1.16	2.1E-01	AJ223392.1	NT	Doto fragilis mitochondrial 16S rRNA gene, partial
6979	19477	32299	2.04	2.1E-01	U04642.1	NT	Human olfactory receptor (OR17-2) gene, partial cds
7436	19960	32825	1.24	2.1E-01	Q01956	SWISSPROT	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.3 (KSHIIID)
7436	19960	32826	1.24	2.1E-01	Q01956	SWISSPROT	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV3.3 (KSHIIID)
7447	19971		2.17	2.1E-01	AE000972.1	NT	Archaeoglobus fulgidus section 135 of 172 of the complete genome
7692	20201	33088	2.02	2.1E-01	AF000949.1	NT	Canis familiaris keratin (KRT9) gene, complete cds
7731	20239	33130	1.14	2.1E-01	AF068687.1	NT	Glycine max malate dehydrogenase (Mdh-2) gene, nuclear gene encoding mitochondrial protein, partial cds
7731	20239	33131	1.14	2.1E-01	AF068687.1	NT	Glycine max malate dehydrogenase (Mdh-2) gene, nuclear gene encoding mitochondrial protein, partial cds
7765	20273		0.68	2.1E-01	T87354.1	EST_HUMAN	yd83b01.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:114793 5'
8017	20559		1.19	2.1E-01	7305030	NT	Mus musculus erythrocyte protein band 4.1-like 3 (Epb4.1l3), mRNA
8439	20979	33894	4.93	2.1E-01	U68399.1	NT	Haemophilus influenzae hmcD, putative haemocin processing protein (hmcC), putative ABC transporter (hmcB), putative haemocin structural protein (hmcA), and haemocin immunity protein (hmcI) genes, complete cds
8732	21271	34190	0.82	2.1E-01	AL040537.1	EST_HUMAN	DKFZp434H0614_r1 434 (synonym: htes3) Homo sapiens cDNA clone DKFZp434H0614 5'
8732	21271	34191	0.82	2.1E-01	AL040537.1	EST_HUMAN	DKFZp434H0614_r1 434 (synonym: htes3) Homo sapiens cDNA clone DKFZp434H0614 5'
8888	21426		0.47	2.1E-01	AB022524.1	NT	Homo sapiens APCL gene, exon 9
8967	21505	34426	5.93	2.1E-01	Z35786.1	NT	S.cerevisiae chromosome II reading frame ORF YBL025w
9423	21932	34880	0.6	2.1E-01	N42536.1	EST_HUMAN	yy11e10.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:270954 5'
9423	21932	34881	0.6	2.1E-01	N42536.1	EST_HUMAN	yy11e10.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:270954 5'
9432	21958	34906	2.95	2.1E-01	X97378.1	NT	A.thaliana mRNA for AtRanBP1b protein
9536	22036	34996	1.57	2.1E-01	AB036529.1	NT	Homo sapiens p53R2 gene for ribonucleotide reductase, exon 6
10232	22727	35719	1.04	2.1E-01	Z97067.1	NT	Beta vulgaris mRNA for elongation factor 1-beta
10263	22758	35745	1.96	2.1E-01	P52824	SWISSPROT	DIACYLGLYCEROL KINASE, DELTA (DIGLYCERIDE KINASE) (DGK-DELTA) (DAG KINASE DELTA) (80 KD DIACYLGLYCEROL KINASE)
10269	22764	35751	0.67	2.1E-01	BF574254.1	EST_HUMAN	602131427F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4270831 5'

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10505	22999	36007	0.5	2.1E-01	AF294296.1	NT	Anolis lineatopus isolate NG NADH dehydrogenase subunit 2 (ND2) gene, complete cds; mitochondrial gene for mitochondrial product
11438	23888		2.24	2.1E-01	11036647	NT	Homo sapiens pancreatic polypeptide 2 (PPY2), mRNA
11451	23901	36969	2.34	2.1E-01	BE180422.1	EST_HUMAN	RC3-HT0622-040500-013-b11 HT0622 Homo sapiens cDNA
11841	24602		1.39	2.1E-01	X57624.1	NT	Drosophila melanogaster ALA-E6 DNA, repeat region
12183	24418		1.46	2.1E-01	AF217490.1	NT	Homo sapiens fragile 16D oxido reductase (FOR) gene, exons 8, 9, and partial cds
12465	24593		1.72	2.1E-01	BE622149.1	EST_HUMAN	601440712F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915675 5'
12607	24691	30858	2.08	2.1E-01	BE672330.1	EST_HUMAN	7a59e02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3223034 3'
12612	24695	30861	1.26	2.1E-01	AJ276505.1	NT	Mus musculus genomic fragment, 279 Kb, chromosome 7
214	12875	25362	1.86	2.0E-01	AB017437.1	NT	Gallus gallus mRNA for avena, complete cds
559	13190		2.2	2.0E-01	7705601	NT	Homo sapiens CGI-18 protein (LOC51008), mRNA
728	13348	25840	1.24	2.0E-01	M77085.1	NT	O.cuniculus germline IgH heavy chain V-H pseudogene, allotype VHa2
843	13459	25968	1.76	2.0E-01	AF027865.1	NT	Mus musculus Major Histocompatibility Locus class II region
1049	13656	26187	0.72	2.0E-01	D90905.1	NT	Synechocystis sp. PCC6803 complete genome, 7/27, 781449-920915
1164	13766	26276	3.24	2.0E-01	AL163213.2	NT	Homo sapiens chromosome 21 segment HS21C013
1297	13891	26414	1.37	2.0E-01	AJ132695.5	NT	Homo sapiens rac1 gene
1351	13946	26470	1.22	2.0E-01	AW384937.1	EST_HUMAN	PM1-HT0422-291299-002-c06 HT0422 Homo sapiens cDNA
1507	14099		1.22	2.0E-01	AJ243957.1	NT	Plum pox virus strain M, complete genome, isolate PS
1534	14126	26663	23.08	2.0E-01	4503408	NT	Homo sapiens dystrobrevin, alpha (DTNA), mRNA
1599	14191	26722	3.03	2.0E-01	AB007974.1	NT	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0505
1604	14196	26726	1.23	2.0E-01	AF260700.1	NT	Homo sapiens sodium/iodide symporter mRNA, partial cds
1735	14326	26868	1.17	2.0E-01	U22348.1	NT	Human bradykinin B1 receptor (bradyb1) gene, complete cds
1755	14345		1.83	2.0E-01	AF111170.3	NT	Homo sapiens 14q32 Jagged2 gene, complete cds; and unknown gene
1795	14385		1.99	2.0E-01	U67525.1	NT	Methanococcus jannaschii section 67 of 150 of the complete genome
1934	14518	27073	1.14	2.0E-01	BE871330.1	EST_HUMAN	601449441F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853330 5'
1934	14518	27074	1.14	2.0E-01	BE871330.1	EST_HUMAN	601449441F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853330 5'
1937	14521	27077	1	2.0E-01	8922238	NT	Homo sapiens hypothetical protein FLJ10120 (FLJ10120), mRNA
2388	14955		1.64	2.0E-01	X82877.1	NT	H.sapiens Na+-D-glucose cotransport regulator gene
2915	15532		0.66	2.0E-01	AF074990.1	NT	Homo sapiens full length insert cDNA YH85A11
3534	16139	28621	0.7	2.0E-01	P46807	SWISSPROT	HOMEOBOX PROTEIN GLABRA2 (HOMEOBOX-LEUCINE ZIPPER PROTEIN ATHB-10) (HD-ZIP PROTEIN ATHB-10)
3626	16229		0.82	2.0E-01	AW238005.1	EST_HUMAN	xp15b02.x1 NCI_CGAP_HN9 Homo sapiens cDNA clone IMAGE:2740395 3' similar to contains element MER21 repetitive element ;
3768	16369	28835	0.8	2.0E-01	P34641	SWISSPROT	CED-11 PROTEIN

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
4028	16626	29098	0.78	2.0E-01	Z46906.1	NT	Sus scrofa
4102	16696	29152	0.68	2.0E-01	X83997.1	NT	C.parasitica eapC gene
4522	17106	29552	0.76	2.0E-01	AF242431.1	NT	Mus musculus neuronal apoptosis inhibitory protein 6 (Naip6) gene, complete cds; and Naip3 gene, exons 2-9 and 11-16
4665	17247		8.43	2.0E-01	BE826165.1	EST_HUMAN	QV4-EN0032-190500-223-e03 EN0032 Homo sapiens cDNA
5192	17757	30186	7.09	2.0E-01	8922080	NT	Homo sapiens hypothetical protein ASH1 (ASH1), mRNA
5226	16139	28621	0.62	2.0E-01	P46607	SWISSPROT	HOMEOBOX PROTEIN GLABRA2 (HOMEOBOX-LEUCINE ZIPPER PROTEIN ATHB-10) (HD-ZIP PROTEIN ATHB-10)
5636	18265	30737	2.38	2.0E-01	X56600.1	NT	Rat SOD-2 gene for manganese-containing superoxide dismutase
5916	18538	31263	2	2.0E-01	11432540	NT	Homo sapiens dual oxidase-like domains 2 (DUOX2), mRNA
6006	18626	31361	0.69	2.0E-01	X91856.1	NT	F.rubripes DNA encoding for valyl-tRNA synthetase
6210	18820	31591	6.48	2.0E-01	U15300.1	NT	Saccharomyces cerevisiae Hal5p (HAL5) mRNA, complete cds
6321	18928		0.71	2.0E-01	M75967.1	NT	Human hepatocyte growth factor gene, exon 1
6560	19158	31955	3.94	2.0E-01	X61033.1	NT	M.auratus mu class glutathione transferase gene
6650	19246	32049	3.63	2.0E-01	AW360865.1	EST_HUMAN	PM1-CT0247-141099-001-g06 CT0247 Homo sapiens cDNA
7251	19780	32636	0.68	2.0E-01	U39724.1	NT	Mycoplasma genitalium section 46 of 51 of the complete genome
7336	19863	32727	1.18	2.0E-01	AF250371.1	NT	Mus musculus phosphofructokinase-1 C isozyme (Pfkc) gene, exons 3 through 7
7775	20284	33181	1.53	2.0E-01	AK024427.1	NT	Homo sapiens mRNA for FLJ00016 protein, partial cds
7895	20437		6.45	2.0E-01	AF028026.1	NT	Andes virus strain OI23133 glycoprotein G1 and G2 precursor, gene, partial cds
8142	20883	33595	2.91	2.0E-01	X91151.1	NT	M.musculus scp2 gene exon 14
8658	21197		0.53	2.0E-01	BE562247.1	EST_HUMAN	601344648F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3677794 5'
9273	21799	34749	1.03	2.0E-01	U82511.1	NT	Dictyostelium discoideum random slug cDNA19 protein (rsc19) mRNA, partial cds
9312	21826	34775	0.65	2.0E-01	U71122.1	NT	Arabidopsis pyruvate decarboxylase-2 (Pdc2) gene, complete cds
9475	21874		4.35	2.0E-01	AE001278.1	NT	Chlamydia trachomatis section 5 of 87 of the complete genome
9661	22160	35132	0.51	2.0E-01	P11420	SWISSPROT	DAUGHTERLESS PROTEIN
9661	22160	35133	0.51	2.0E-01	P11420	SWISSPROT	DAUGHTERLESS PROTEIN
9806	22304		1.98	2.0E-01	AF146692.1	NT	Homo sapiens filamin 2 (FLN2) mRNA, complete cds
9954	22449	35431	1.79	2.0E-01	AF086907.1	NT	Arabidopsis thaliana root gravitropism control protein (PIN2) gene, complete cds
9954	22449	35432	1.79	2.0E-01	AF086907.1	NT	Arabidopsis thaliana root gravitropism control protein (PIN2) gene, complete cds
10072	22567	35562	0.53	2.0E-01	AF157814.1	NT	Homo sapiens cAMP specific phosphodiesterase (PDE4C) gene, exons 2 through 12
10072	22567	35563	0.53	2.0E-01	AF157814.1	NT	Homo sapiens cAMP specific phosphodiesterase (PDE4C) gene, exons 2 through 12
10115	22610		0.72	2.0E-01	X78388.1	NT	D.melanogaster DNA mobile element (hoppel)
10304	22798	35789	0.88	2.0E-01	X97121.1	NT	R.norvegicus mRNA for NTR2 receptor
10720	23248	36263	2.77	2.0E-01	D89088.1	NT	Salvelinus pluvius mRNA for transferrin, complete cds

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10720	23248	36264	2.77	2.0E-01	D89088.1	NT	Salvelinus pluvius mRNA for transferrin, complete cds
12162	24402		1.34	2.0E-01	AF206637.2	NT	Pimephales promelas liver glucose-6-phosphate-1-dehydrogenase mRNA, partial cds
12374	24887		1.95	2.0E-01	AF302773.1	NT	Homo sapiens ninein-Lm isoform (ninein) mRNA, complete cds
12386	24807	30788	2.81	2.0E-01	AW975297.1	EST_HUMAN	EST387405 MAGE resequences, MAGN Homo sapiens cDNA
12425	24810	30888	3.97	2.0E-01	AI023592.1	EST_HUMAN	ov80a10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1643610 3'
12449	24584		17.06	2.0E-01	AF078164.2	NT	Homo sapiens Ku70-binding protein (KUB3) mRNA, partial cds
115	12786		6.22	1.9E-01	7549743	NT	Rattus norvegicus Aryl hydrocarbon receptor nuclear translocator 1 (Arnt1), mRNA
374	13023	25509	5.4	1.9E-01	AF004353.1	NT	Mus musculus pale ear (ep) gene, wild type allele, 3' region, partial cds
684	13308	25792	1.47	1.9E-01	U32581.2	NT	Homo sapiens lambda/iota protein kinase C-interacting protein mRNA, complete cds
684	13308	25793	1.47	1.9E-01	U32581.2	NT	Homo sapiens lambda/iota protein kinase C-interacting protein mRNA, complete cds
691	13315	25800	6.6	1.9E-01	BE070801.1	EST_HUMAN	RC3-BT0502-251199-011-d01 BT0502 Homo sapiens cDNA
692	13315	25800	6.82	1.9E-01	BE070801.1	EST_HUMAN	RC3-BT0502-251199-011-d01 BT0502 Homo sapiens cDNA
1023	13633		1.92	1.9E-01	7305180	NT	Mus musculus interleukin 2 receptor, gamma chain (Il2rg), mRNA
1143	13746	26256	10.04	1.9E-01	AA358813.1	EST_HUMAN	EST67784 Fetal lung II Homo sapiens cDNA 5' end
1413	14008	26534	2.41	1.9E-01	AF061282.1	NT	Sorghum bicolor 22 kDa kafirin cluster
1482	14075		4.02	1.9E-01	AF184623.1	NT	Plasmodium vivax reticulocyte binding protein-2 (rbp-2) gene, complete cds
2185	14761	27331	1.29	1.9E-01	AA916492.1	EST_HUMAN	ol44h09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1526369 3' similar to gb:A03911 GLIA DERIVED NEXIN PRECURSOR (HUMAN);
2422	14990	27563	3.27	1.9E-01	8922533	NT	Homo sapiens hypothetical protein FLJ10581 (FLJ10581), mRNA
2949	15565	28039	4.1	1.9E-01	U66066.1	NT	Sigmodon hispidus p53 gene, partial cds
2965	15580		6.58	1.9E-01	J00922.1	NT	Gallus gallus ovalbumin (Y) gene, complete cds
3033	15649	28128	1.05	1.9E-01	U25148.1	NT	Rattus norvegicus brush border myosin-I (BBMI) mRNA, partial cds
3442	16050	28528	4.19	1.9E-01	D13197.1	NT	Mouse gene for immunoglobulin diversity region D1
3526	16131	28611	5.24	1.9E-01	R16467.1	EST_HUMAN	yf42f10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:129547 5'
3877	16475	28939	0.76	1.9E-01	AF264017.1	NT	Rattus norvegicus arylacetamide deacetylase gene, complete cds
4064	16661	29123	3.85	1.9E-01	AB006784.1	NT	Schizosaccharomyces pombe DNA for cytoplasmic dynein heavy chain, complete cds
4157	16749	29202	1.86	1.9E-01	AW754106.1	EST_HUMAN	CM3-CT0315-271199-045-b11 CT0315 Homo sapiens cDNA
4315	16901	29345	1.17	1.9E-01	BE834943.1	EST_HUMAN	MR1-FN0010-290700-007-d04 FN0010 Homo sapiens cDNA
4568	17151	29597	0.69	1.9E-01	AL161493.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 5
5156	17726		1.11	1.9E-01	AF223642.1	NT	Rattus norvegicus chemokine receptor CXCR3 mRNA, complete cds
5789	18414		5.46	1.9E-01	AW130149.1	EST_HUMAN	xf29a07.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2619444 3' similar to gb:M73779 RETINOIC ACID RECEPTOR ALPHA-1 (HUMAN);
5826	18450	31173	7.81	1.9E-01	AF127937.1	NT	Homo sapiens DNA polymerase epsilon catalytic subunit protein (POLE1) gene, exon 1a
6005	18625	31360	0.73	1.9E-01	AF091216.1	NT	Mus musculus Wrn protein (Wrn) gene, complete cds

WO 01/57277 PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6046	18665		2.52	1.9E-01	AU133116.1	EST_HUMAN	AU133116 NT2RP4 Homo sapiens cDNA clone NT2RP4001328 5'
6469	19070	31855	1.07	1.9E-01	AI762391.1	EST_HUMAN	wi54h02.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2394099 3'
6523	19123	31915	1.23	1.9E-01	AW148452.1	EST_HUMAN	xf14c08.x1 NCI_CGAP_Kid8 Homo sapiens cDNA clone IMAGE:2618030 3' similar to gb:X03559 ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR (HUMAN);
7050	18069	30460	1.37	1.9E-01	R43212.1	EST_HUMAN	yg09e12.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:31663 3' similar to contains MER13 repetitive element ;
7072	19644	32481	0.91	1.9E-01	AF034920.1	NT	Homo sapiens tubby like protein 1 (TULP1) gene, exons 9-11
7072	19644	32482	0.91	1.9E-01	AF034920.1	NT	Homo sapiens tubby like protein 1 (TULP1) gene, exons 9-11
7503	20025	32889	1.3	1.9E-01	U80922.1	NT	Arabidopsis thaliana serine/threonine protein phosphatase type one (TOPP8) gene, complete cds
7543	20083	32937	2.89	1.9E-01	AF072724.1	NT	Zea mays starch branching enzyme I (sbe1) gene, complete cds
7927	20469	33378	1.71	1.9E-01	AL161557.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57
8620	21159	34074	12.12	1.9E-01	AB033024.1	NT	Homo sapiens mRNA for KIAA1198 protein, partial cds
8875	21414	34337	1.36	1.9E-01	M14568.1	NT	Marsupial cat beta-globin gene mRNA, partial cds
8875	21414	34338	1.36	1.9E-01	M14568.1	NT	Marsupial cat beta-globin gene mRNA, partial cds
9789	22287	35271	0.72	1.9E-01	AA912486.1	EST_HUMAN	ol96g10.s1 NCI_CGAP_PNS1 Homo sapiens cDNA clone IMAGE:1537506 3' similar to contains Alu repetitive element;
10140	22635	35626	0.71	1.9E-01	BE830353.1	EST_HUMAN	RC5-ET0082-060700-022-A02 ET0082 Homo sapiens cDNA
10140	22635	35627	0.71	1.9E-01	BE830353.1	EST_HUMAN	RC5-ET0082-060700-022-A02 ET0082 Homo sapiens cDNA
10523	23061	36071	2.02	1.9E-01	AL161503.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 15
10523	23061	36072	2.02	1.9E-01	AL161503.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 15
10635	23167	36178	2.06	1.9E-01	AF223391.1	NT	Homo sapiens calcium channel alpha1E subunit (CACNA1E) gene, exons 7-49, and partial cds, alternatively spliced
11377	23829	36891	1.68	1.9E-01	M22253.1	NT	Rattus norvegicus sodium channel I mRNA, complete cds
11571	24018	37088	2.69	1.9E-01	AJ243213.1	NT	Homo sapiens partial 5-HT4 receptor gene, exons 2 to 5
12207	24431		1.33	1.9E-01	AF055900.1	NT	Drosophila melanogaster clathrin light chain mRNA, complete cds
12582	24874		3.69	1.9E-01	AF001168.1	NT	Arabidopsis thaliana receptor-like kinase LECRK1 (LECRK1) gene, complete cds
34	12713	25172	2.56	1.8E-01	U73200.1	NT	Mus musculus p116Rip mRNA, complete cds
281	15412	25423	1.67	1.8E-01	AB022090.1	NT	Mus musculus Cctg gene for chaperonin containing TCP-1 gamma subunit, partial cds
393	13039	25530	1.76	1.8E-01	4502532	NT	Homo sapiens calcium channel, voltage-dependent, beta 2 subunit (CACNB2) mRNA, and translated products
776	13395	25896	0.77	1.8E-01	AB021490.2	NT	Oryzias latipes gene for membrane guanylyl cyclase OlGC1, complete cds
1018	13628	26141	0.78	1.8E-01	AI912212.1	EST_HUMAN	wd71f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2337051 3'
1130	13732	26242	1.26	1.8E-01	AF000580.1	NT	Dictyostelium discoideum plasmid Ddp5, complete genome
1332	13926	26447	6.97	1.8E-01	AL117189.1	NT	Yersinia pestis plasmid pCD1

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1551	14143	26676	1.31	1.8E-01	6753947	NT	Mus musculus guanylate nucleotide binding protein 1 (Gbp1), mRNA
1551	14143	26677	1.31	1.8E-01	6753947	NT	Mus musculus guanylate nucleotide binding protein 1 (Gbp1), mRNA
1887	14472		2.79	1.8E-01	4505036	NT	Homo sapiens latent transforming growth factor beta binding protein 4 (LTBP4) mRNA
1907	14492		2.22	1.8E-01	AI733708.1	EST_HUMAN	qg22d10.x5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1761811 3' similar to TR:O75936 O75936 GAMMA BUTYROBETAINE HYDROXYLASE ;
1958	14542	27098	1.52	1.8E-01	AB051897.1	NT	Mus musculus Scya6, Scya9, Scya16-ps, Scya5 genes for small inducible cytokine A6 precursor, small inducible cytokine A9 precursor, Scya16 pseudogene, small inducible cytokine A5 precursor, complete cds
2716	15273		2.29	1.8E-01	AW935728.1	EST_HUMAN	QV3-DT0018-081299-036-g04 DT0018 Homo sapiens cDNA
2923	15540		2.36	1.8E-01	AF184589.1	NT	Jonopsidium acaule LEAFY protein (LEAFY2) gene, partial cds
2928	15544	28020	1.18	1.8E-01	AW182300.1	EST_HUMAN	yj41a03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2659756 3'
3158	15772	28239	1.31	1.8E-01	AW995178.1	EST_HUMAN	QV0-BN0041-070300-147-c04 BN0041 Homo sapiens cDNA
3413	16021	28501	0.71	1.8E-01	BF183582.1	EST_HUMAN	601809723R1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040621 3'
3683	16284	28752	0.79	1.8E-01	H03369.1	EST_HUMAN	yj45e01.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:151704 3' similar to contains Alu repetitive element;
3683	16284	28753	0.79	1.8E-01	H03369.1	EST_HUMAN	yj45e01.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:151704 3' similar to contains Alu repetitive element;
4333	16920	29362	0.78	1.8E-01	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
4426	17012		4.07	1.8E-01	D37954.1	NT	Bovine NB25 mRNA for MHC class II (BoLA-DQB), complete cds
4654	17236	29691	6.59	1.8E-01	AL161556.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 56
4886	17461	29914	2.51	1.8E-01	AB051897.1	NT	Mus musculus Scya6, Scya9, Scya16-ps, Scya5 genes for small inducible cytokine A6 precursor, small inducible cytokine A9 precursor, Scya16 pseudogene, small inducible cytokine A5 precursor, complete cds
4928	17503	29950	1.03	1.8E-01	X92179.1	NT	S.tuberosum mRNA for alcohol dehydrogenase
5198	17763	30188	2.18	1.8E-01	AW814270.1	EST_HUMAN	MR3-ST0203-151299-112-g06 ST0203 Homo sapiens cDNA
5216	17781	30200	1.59	1.8E-01	AI792382.1	EST_HUMAN	an28g07.y5 Gessler Wilms tumor Homo sapiens cDNA clone IMAGE:1700028 5'
5257	17820	30245	1.5	1.8E-01	AF181258.1	NT	Mesocricetus auratus Na-taurocholate cotransporting polypeptide mRNA, partial cds
5281	17843	30270	1.07	1.8E-01	AI439881.1	EST_HUMAN	tt57e04.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2134590 3'
5288	17850	30276	0.59	1.8E-01	AF132115.1	NT	Arabidopsis thaliana cytochrome b-561 (CYTB561) gene, partial cds
5338	17899	30314	0.78	1.8E-01	AJ132844.1	NT	Broad bean wilt virus 2 genes encoding 119kDa protein, 104kDa protein, large coat protein, small coat protein
5338	17899	30315	0.78	1.8E-01	AJ132844.1	NT	Broad bean wilt virus 2 genes encoding 119kDa protein, 104kDa protein, large coat protein, small coat protein
5398	17956	30367	2.04	1.8E-01	AW809402.1	EST_HUMAN	MR4-ST0121-041199-019-b01 ST0121 Homo sapiens cDNA

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5976	18596	31331	1	1.8E-01	AL161594.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 90
6082	18699	31446	1.01	1.8E-01	N28629.1	EST_HUMAN	yx38h08.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:264063 5'
6277	18885	31653	1.1	1.8E-01	6678428	NT	Mus musculus Tnf receptor-associated factor 6 (Traf6), mRNA
6277	18885	31654	1.1	1.8E-01	6678428	NT	Mus musculus Tnf receptor-associated factor 6 (Traf6), mRNA
6635	19231	32035	2.03	1.8E-01	Q9QY14	SWISSPROT	FORKHEAD BOX PROTEIN E3
6675	19271		2.24	1.8E-01	N94853.1	EST_HUMAN	yy62h02.r1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:278163 5'
7077	19649	32487	1.22	1.8E-01	AB018561.1	NT	Citrullus lanatus mRNA for wsus, complete cds
7077	19649	32488	1.22	1.8E-01	AB018561.1	NT	Citrullus lanatus mRNA for wsus, complete cds
7117	19457	32272	0.71	1.8E-01	BE961353.1	EST_HUMAN	601648361R2 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:3932247 3'
8547	21086	34009	0.47	1.8E-01	AW966118.1	EST_HUMAN	EST378191 MAGE resequences, MAGI Homo sapiens cDNA
9266	21792	34741	1.13	1.8E-01	M73258.1	NT	Human cellular DNA/Human papillomavirus proviral DNA
9296	21896	34843	1.39	1.8E-01	9626232	NT	Bacteriophage lke, complete genome
9412	21921		0.55	1.8E-01	AA493751.1	EST_HUMAN	nh02a05.s1 NCI_CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943088 similar to contains L1.t3 L1 repetitive element ;
9494	21994	34950	1.13	1.8E-01	P15272	SWISSPROT	AMP NUCLEOSIDASE
9494	21994	34951	1.13	1.8E-01	P15272	SWISSPROT	AMP NUCLEOSIDASE
9532	22032	34990	0.95	1.8E-01	M26019.1	NT	S.commune orotidine-5'-phosphate decarboxylase (URA1) gene, complete cds
9532	22032	34991	0.95	1.8E-01	M26019.1	NT	S.commune orotidine-5'-phosphate decarboxylase (URA1) gene, complete cds
9694	22193	35166	0.62	1.8E-01	P08123	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
9698	22197	35170	0.69	1.8E-01	U67548.1	NT	Methanococcus jannaschii section 90 of 150 of the complete genome
10039	22534		0.64	1.8E-01	AF200252.1	NT	Aquarius amplus cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial gene for mitochondrial product
10271	22766	35753	1.22	1.8E-01	X63440.1	NT	M.musculus mRNA for P19-protein tyrosine phosphatase
10518	23054	36066	2.37	1.8E-01	X77336.1	NT	A.thaliana mRNA for ribonucleotide reductase R2
10558	23094	36106	7.47	1.8E-01	U38906.1	NT	Bacteriophage r11 integrase, repressor protein (rro), dUTPase, holin and lysin genes, complete cds
10615	19649	32487	3.07	1.8E-01	AB018561.1	NT	Citrullus lanatus mRNA for wsus, complete cds
10615	19649	32488	3.07	1.8E-01	AB018561.1	NT	Citrullus lanatus mRNA for wsus, complete cds
10616	23148	36160	4.49	1.8E-01	AF019107.1	NT	Dictyostelium discoideum unknown (DG1041) gene, complete cds
10897	23417	36434	1.84	1.8E-01	M59257.1	NT	Human carcinoembryonic antigen (CEA) gene, exon 4
11337	23035	36045	4.3	1.8E-01	X57033.1	NT	B.taurus mRNA for potassium channel
11599	24042	37111	2.74	1.8E-01	8394421	NT	Rattus norvegicus Thromboxane receptor (Tbxa2r), mRNA
11626	24068	37132	1.6	1.8E-01	U40487.1	NT	Mycobacterium smegmatis proton antiporter efflux pump (lfrA), complete cds
11748	24146		2.04	1.8E-01	10086561	NT	Bovine ephemeral fever virus, complete genome

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11814	24186	31032	1.41	1.8E-01	BF348623.1	EST_HUMAN	602019928F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4155318 5'
12218	13926	26447	1.3	1.8E-01	AL117189.1	NT	Yersinia pestis plasmid pCD1
12291	24491		5.61	1.8E-01	Q96682	SWISSPROT	DNA TERMINAL PROTEIN (BELLETT PROTEIN) (PTP PROTEIN)
12416	24569		23.47	1.8E-01	R24494.1	EST_HUMAN	yh48h10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:133027 5'
12459	24590		2.75	1.8E-01	Y11114.1	NT	E.dispar mRNA for hexokinase (hxk1)
12502	25045	30507	1.58	1.8E-01	9506952	NT	Rattus norvegicus procollagen C-proteinase enhancer protein (Pcolce), mRNA
603	13232	25705	5.53	1.7E-01	BE385164.1	EST_HUMAN	601274604F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3615768 5'
838	13454	25964	2.99	1.7E-01	X53330.1	NT	P.dumerilii histone gene cluster for core histones H2A, H2B, H3 and H4
996	13608		8.63	1.7E-01	P35616	SWISSPROT	NEUROFILAMENT TRIPLET L PROTEIN (NEUROFILAMENT LIGHT POLYPEPTIDE) (NF-L)
1096	13701	26210	0.67	1.7E-01	AF081810.1	NT	Lymantria dispar nucleopolyhedrovirus, complete genome
1096	13701	26211	0.67	1.7E-01	AF081810.1	NT	Lymantria dispar nucleopolyhedrovirus, complete genome
1853	14441	26998	0.95	1.7E-01	AL161573.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 69
2025	14607		2.84	1.7E-01	AF255051.1	NT	Homo sapiens BNIP3H (BNIP3H) gene, complete cds; nuclear gene for mitochondrial product
2885	15503	27973	1.98	1.7E-01	AF000716.1	NT	Vibrio cholerae hypoxanthine phosphoribosyltransferase (hpt) gene, partial cds, hemagglutinin/protease regulatory protein (hapR) gene, complete cds, and YRAL VIBCO gene, partial cds
2885	15503	27974	1.98	1.7E-01	AF000716.1	NT	Vibrio cholerae hypoxanthine phosphoribosyltransferase (hpt) gene, partial cds, hemagglutinin/protease regulatory protein (hapR) gene, complete cds, and YRAL VIBCO gene, partial cds
2953	15569	28044	1.53	1.7E-01	AA336909.1	EST_HUMAN	EST41651 Endometrial tumor Homo sapiens cDNA 5' end
3027	15643	28121	1.9	1.7E-01	AJ238736.1	NT	Naja naja atra cbx-1 gene, exons 1-3
3027	15643	28122	1.9	1.7E-01	AJ238736.1	NT	Naja naja atra cbx-1 gene, exons 1-3
3139	15753	28220	1.91	1.7E-01	AF081514.1	NT	Taxus canadensis geranylgeranyl diphosphate synthase mRNA, complete cds
3412	16020	28500	1.11	1.7E-01	N55763.1	EST_HUMAN	J2346F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone J2346 5'
3494	16099	28574	1.26	1.7E-01	AJ269505.1	NT	Anabaena sp. ORF4 (partial), ORF3, ORF2, ORF1, adpA gene, adpB gene, adpC gene, adpD gene, adpE gene and adpF gene
4012	16610	29083	4.99	1.7E-01	AJ235377.1	NT	Homo sapiens derivative 11 breakpoint fragment: partial intron 10 of the ALL-1/MLL/HRX gene fused to intron 5 of the AF-4/FEL gene
4658	17240		1.63	1.7E-01	X52936.1	NT	Schistocerca gregaria alpha repetitive DNA
4877	17452	29904	0.84	1.7E-01	AF217490.1	NT	Homo sapiens fragile 16D oxido reductase (FOR) gene, exons 8, 9, and partial cds
4965	17539	29981	1.07	1.7E-01	AI247635.1	EST_HUMAN	qh57e09.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1848808 3' similar to contains OFR.b1 OFR repetitive element ;
5210	17775		0.88	1.7E-01	U28376.1	NT	Zea mays calcium-dependent protein kinase (MZECDPK2) mRNA, complete cds
5242	17806	30227	1.02	1.7E-01	BF689719.1	EST_HUMAN	602186630F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4298646 5'

WO 01/57277

PCT/US01/00669

Table 4

Single Exon Probes Expressed in Fetal Liver

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5297	17859		1.08	1.7E-01	AF072725.1	NT	Zea mays starch branching enzyme IIb (se) gene, complete cds
5342	17903	30319	0.6	1.7E-01	BF030010.1	EST_HUMAN	601557256F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3827197 5'
5421	17978	30386	7.82	1.7E-01	J04479.1	NT	S.pneumoniae DNA polymerase I (polA) gene, complete cds
5604	18233	30683	1.92	1.7E-01	AA470686.1	EST_HUMAN	ne13a02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:881066 3' similar to gb:M17886 60S ACIDIC RIBOSOMAL PROTEIN P1 (HUMAN);
5604	18233	30684	1.92	1.7E-01	AA470686.1	EST_HUMAN	ne13a02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:881066 3' similar to gb:M17886 60S ACIDIC RIBOSOMAL PROTEIN P1 (HUMAN);
5779	18404	31120	0.7	1.7E-01	U43599.1	NT	Brugia pahangi microfilarial sheath protein SHP3 (shp3) gene, complete cds
6471	19072	31856	20.9	1.7E-01	H72118.1	EST_HUMAN	ys02g06.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:213658 3'
6522	19122	31913	1.33	1.7E-01	AI370976.1	EST_HUMAN	ta29c11.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:2045492 3'
6522	19122	31914	1.33	1.7E-01	AI370976.1	EST_HUMAN	ta29c11.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:2045492 3'
6937	18045	30467	0.71	1.7E-01	BE300286.1	EST_HUMAN	600944067T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960248 3'
6960	19537		2.26	1.7E-01	AF026552.3	NT	Mesocricetus auratus oviductin precursor (OVI) gene, complete cds
7074	19646		0.67	1.7E-01	Z92910.1	NT	Homo sapiens HFE gene
7272	19800	32657	2.83	1.7E-01	AP000422.1	NT	Escherichia coli O157:H7 genomic DNA, Sakai-VT2 prophage inserted region
7339	19866	32730	8.92	1.7E-01	BE734179.1	EST_HUMAN	601569022F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3843964 5'
7494	20017	32882	1.16	1.7E-01	P16724	SWISSPROT	PROBABLE PROCESSING AND TRANSPORT PROTEIN UL56 (HFLF0 PROTEIN)
7508	24784	32893	0.73	1.7E-01	Q01955	SWISSPROT	COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR
7802	20345	33253	1.26	1.7E-01	AF000573.1	NT	Homo sapiens homogentisate 1,2-dioxygenase gene, complete cds
7904	20446	33352	0.54	1.7E-01	AF150669.1	NT	Pseudomonas putida long-chain-fatty-acid-CoA ligase (fadD) gene, complete cds
8219	20760	33674	6.62	1.7E-01	7706426	NT	Homo sapiens cleavage and polyadenylation specificity factor 3, 73kD subunit (CPSF3), mRNA
8219	20760	33675	6.62	1.7E-01	7706426	NT	Homo sapiens cleavage and polyadenylation specificity factor 3, 73kD subunit (CPSF3), mRNA
8631	21170	34087	0.58	1.7E-01	AW992873.1	EST_HUMAN	RC2-BN0032-120200-011-a10 BN0032 Homo sapiens cDNA
8662	21201	34119	3.28	1.7E-01	D00384.1	NT	Rat (SHR strain) SX1 gene
8778	21317	34239	0.68	1.7E-01	AF217413.1	NT	Homo sapiens neuroligin 3 isoform gene, complete cds, alternatively spliced
8778	21317	34240	0.68	1.7E-01	AF217413.1	NT	Homo sapiens neuroligin 3 isoform gene, complete cds, alternatively spliced
9095	21631	34569	0.46	1.7E-01	BE253142.1	EST_HUMAN	601116672F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3357184 5'
9095	21631	34570	0.46	1.7E-01	BE253142.1	EST_HUMAN	601116672F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3357184 5'
9509	22009	34967	7.72	1.7E-01	AP001508.1	NT	Bacillus halodurans genomic DNA, section 2/14
9614	22114	35077	0.56	1.7E-01	AW977455.1	EST_HUMAN	EST389564 MAGE resequences, MAGO Homo sapiens cDNA
9614	22114	35078	0.56	1.7E-01	AW977455.1	EST_HUMAN	EST389564 MAGE resequences, MAGO Homo sapiens cDNA
9631	22131	35096	2.47	1.7E-01	U16288.1	NT	Human class IV alcohol dehydrogenase (ADH7) gene, exon 3
9704	22203	35174	1.27	1.7E-01	Z34508.1	NT	Human immunodeficiency virus type 1 (B7.05) env gene (partial)
9704	22203	35175	1.27	1.7E-01	Z34508.1	NT	Human immunodeficiency virus type 1 (B7.05) env gene (partial)

WO 01/57277

PCT/US01/00669